

- (1) GENERAL INFORMATION
- (i) APPLICANT: Miller, Samuel I.
- (ii) TITLE OF THE INVENTION: SALMONELLA SECRETED PROTEINS AND USES THEREOF
- (iii) NUMBER OF SEQUENCES: 47
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Fish & Richardson, P.C.
  - (B) STREET: 225 Franklin Street
  - (C) CITY: Boston
  - (D) STATE: MA
  - (E) COUNTRY: US
  - (F) ZIP: 02110-2804
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Diskette
  - (B) COMPUTER: IBM Compatible
  - (C) OPERATING SYSTEM: Windows95
  - (D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: 09/068,804
  - (B) FILING DATE: 14-MAY-1998
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: PCT/US96/18504
  - (B) FILING DATE: 14-NOV-1996
  - (A) APPLICATION NUMBER: 60/006,733
  - (B) FILING DATE: 14-NOV-1995
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Fraser, Janis K.
  - (B) REGISTRATION NUMBER: 34,819
  - (C) REFERENCE/DOCKET NUMBER: 00786/292002
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 617-542-5070
  - (B) TELEFAX: 617-542-8906
  - (2) INFORMATION FOR SEQ ID NO:1:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 870 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGCAAAGCCG AGGAAACGAA CCGCATTATG GGATGTATCG GGAAAGTCCT CGGCGCGCTG 60
CTAACCATTG TCAGCGTTGT GGCCGCTGTT TTTACCGGTG GGGCGAGTCT GGCGCTGGCT 120
GCGGTGGGAC TTGCGGTAAT GGTGGCCGAT GAAATTGTGA AGGCGGCGAC GGGAGTGTCG 180

JAN 2 6 1999
MATRIX GULLENTER
SERVICE CENTER

TTTATTCAGC AGGCGCTAAA C ATTGGCAAGG CGATTACCAA A ATGGCCGGCA GCATTGTTGG T GTGGTCGCAG TTGTCGGGAA A ATGGGCGAAA CGATTAAGAA G	AGCGCTGGAA GGA CGCGATTGTC GCC AGGCGCGGCG GCG	ATTAGGCG TCGATAAGAA CGCTATTG CCATGGTGGC GAAACTGG GTAACGCGCT	AACGGCAGAG GGTCATTGTG GAGCAAAATG	240 300 360 420 480
ATGGGCGAAA CGATTAAGAA G AGCAAACTCT TTACCCAGGG G AAGATGGGCC TGCAAACGAA T GTGGCGTTGG GCATGGAAGT C GGCGTATTTA TTAAAAATGC C ATGGATCAGA TTCAGCAGTG G GTAACGGCGG AACTGCAAAA A CGTTTTATTC TGCGCCAGAG T	SATGCAACGT ATT TGCCTTAAGT AAA CACGAATACC GCA CAGCGAGGCG CTT GCTTAAACAA TCC AGCCATGTCT TCT	FACTAGCG GTCTGGGTAA AGAGCTGG TAGGTAATAC AGCCCAGT CAGCCGGTGG FGCTGATT TTATGCTCGC CGTAGAAA TATTTGGTGA	TGTGGGTAGC CCTAAATAAA TGTTGCCGAG CCGTTTTGCC AAACCAGAAG	480 540 600 660 720 780 840 870

# (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1230 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGTTAATTA	GTAATGTGGG	AATAAATCCC	GCCGCTTATT	TAAATAATCA	TTCTGTTGAG	60
AATAGTTCAC	AGACAGCTTC	GCAATCCGTT	AGCGCTAAAG	ATATTCTGAA	TAGTATTGGT	120
ATTAGCAGCA	GTAAAGTCAG	TGACCTGGGG	TTGAGTCCTA	CACTGAGCGC	GCCTGCGCCA	180
GGGGTATTAA	CGCAAACCCC	CGGAACGATC	ACGTCCTTTT	TAAAAGCCAG	TATTCAAAAT	240
ACCGACATGA	ATCAGGATTT	GAATGCTCTG	GCAAATAATG	TCACGACTAA	AGCGAATGAG	300
GTTGTGCAAA	CCCAGTTACG	CGAGCAGCAG	GCAGAAGTCG	GAAAGTTTTT	TGATATTAGC	360
GGAATGTCTT	CCAGTGCCGT	TGCGCTGTTG	GCTGCCGCGA	ATACGTTAAT	GCTGACGTTG	420
AACCAGGCTG	ATAGCAAACT	GTCTGGTAAG	TTGTCATTAG	TCAGTTTTGA	TGCAGCTAAA	480
ACGACGGCAA	GCTCCATGAT	GCGCGAAGGG	ATGAATGCGT	TGTCCGGTAG	TATTTCCCAG	540
AGCGCGCTTC	AGTTGGGGAT	CACTGGCGTG	GGCGCCAAAC	TGGAATATAA	GGGGCTGCAG	600
AATGAAAGAG	GCGCGCTTAA	ACATAATGCC	GCGAAGATCG	ATAAACTGAC	CACTGAAAGC	660
CACAGTATTA	AAAACGTGCT	GAACGGGCAG	AATAGCGTCA	AACTCGGTGC	TGAAGGCGTC	720
GATTCTCTGA	AATCGTTAAA	TATGAAGAAA	ACCGGTACCG	ATGCGACGAA	AAATCTTAAT	780
GATGCGACGC	TTAAATCTAA	TGCCGGAACC	AGCGCCACGG	AAAGTCTGGG	TATTAAAGAC	840
AGTAATAAAC	AAATCTCCCC	TGAACATCAG	GCTATTCTGT	CGAAACGTCT	TGAGTCTGTC	900
GAATCCGATA	TTCGTCTTGA	GCAGAATACC	ATGGATATGA	CCCGAATCGA	TGCGCGCAAG	960
ATGCAGATGA	CGGGCGATCT	GATTATGAAG	AACTCGGTCA	CGGTCGGTGG	TATTGCAGGG	1020
GCGTCCGGGC	AGTACGCCGC	TACTCAGGAA	CGTTCCGAGC	AGCAAATTAG	CCAGGTGAAT	1080
AACCGGGTTG	CCAGCACCGC	ATCGGACGAA	GCCCGTGAAA	GTTCACGTAA	ATCGACCAGC	1140
CTGATTCAGG	AAATGCTGAA	AACAATGGAG	AGCATTAACC	AGTCGAAAGC	ATCCGCACTC	1200
GCTGCTATCG	CAGGCAATAT	TCGCGCTTAA				1230

# (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1032 base pairs

  - (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGCTTAATA	TTCAAAATTA	TTCCGCTTCT	CCTCATCCGG	GGATCGTTGC	CGAACGGCCG	60
CAGACTCCCT	CGGCGAGCGA	GCACGTCGAG	ACTGCCGTGG	TACCGTCTAC	CACAGAACAT	120
CGCGGTACAG	ATATCATTTC	ATTATCGCAG	GCGGCTACTA	AAATCCACCA	GGCACAGCAG	180

ACGCTGCAGT CAA	CGCCACC GATCTCTG	AA GAGAATAATG	ACGAGCGCAC	GCTGGCGCGC	240
CAGCAGTTGA CCA	GCAGCCT GAATGCGC	TG GCGAAGTCCG	GCGTGTCATT	ATCCGCAGAA	300
CAAAATGAGA ACC	TGCGGAG CGCGTTTT	CT GCGCCGACGT	CGGCCTTATT	TAGCGCTTCG	360
CCTATGGCGC AGC	CGAGAAC AACCATTT	CT GATGCTGAGA	TTTGGGATAT	GGTTTCCCAA	420
AATATATCGG CGA	TAGGTGA CAGCTATO	TG GGCGTTTATG	AAAACGTTGT	CGCAGTCTAT	480
ACCGATTTTT ATC	AGGCCTT CAGTGATA	TT CTTTCCAAAA	TGGGAGGCTG	GTTATTACCA	540
GGTAAGGACG GTA	ATACCGT TAAGCTAG	AT GTTACCTCAC	TCAAAAATGA	TTTAAACAGT	600
TTAGTCAATA AAT	ATAATCA AATAAACA	GT AATACCGTTT	TATTTCCAGC	GCAGTCAGGC	660
AGCGGCGTTA AAG	TAGCCAC TGAAGCGG	AA GCGAGACAGT	GGCTCAGTGA	ATTGAATTTA	720
CCGAATAGCT GCC	TGAAATC TTATGGAT	CC GGTTATGTCG	TCACCGTTGA	TCTGACGCCA	780
TTACAAAAAA TGG	TTCAGGA TATTGATG	GT TTAGGCGCGC	CGGGAAAAGA	CTCAAAACTC	840
GAAATGGATA ACG	CCAAATA TCAAGCCT	GG CAGTCGGGTT	TTAAAGCGCA	GGAAGAAAAT	900
ATGAAAACCA CAT	TACAGAC GCTGACGC	AA AAATATAGCA	ATGCCAATTC	ATTGTACGAC	960
AACCTGGTAA AAG	TGCTGAG CAGTACGA	TA AGTAGCAGCC	TGGAAACCGC	CAAAAGCTTC	1020
CTGCAAGGAT AA					1032

# (2) INFORMATION FOR SEQ ID NO:4:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 194 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATGGTTACAA GTGTAAGGAC	TCAGCCCCCC	GTCATAATGC	CAGGTATGCA	GACCGAGATC	60
AAAACGCAGG CCACGAATCT	TGCGGCGAAT	CTTTCCGCAG	TCAGAGAAAG	TGCCACAGCG	120
ACGCTGTCAG GGGAAATTAA	AGGCCCGCAA	CTGGAAGATT	TTCCCGCGCT	GATCAAACAG	180
GCGAGTCTGG ATGC					194

### (2) INFORMATION FOR SEQ ID NO:5:

# (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 290 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: peptide

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Arg Lys Ala Glu Glu Thr Asn Arg Ile Met Gly Cys Ile Gly Lys Val 10 Leu Gly Ala Leu Leu Thr Ile Val Ser Val Val Ala Ala Val Phe Thr 25 20 Gly Gly Ala Ser Leu Ala Leu Ala Ala Val Gly Leu Ala Val Met Val 35 40 Ala Asp Glu Ile Val Lys Ala Ala Thr Gly Val Ser Phe Ile Gln Gln 55 60 Ala Leu Asn Pro Ile Met Glu His Val Leu Lys Pro Leu Met Glu Leu 75 70 Ile Gly Lys Ala Ile Thr Lys Ala Leu Glu Gly Leu Gly Val Asp Arg 90 95 Lys Arg Gln Arg Trp Pro Ala Ala Leu Leu Val Arg Leu Ser Pro Leu 100 105 110 Cys His Gly Asp Ala Val Ile Val Val Val Ala Val Val Gly Lys Gly 120 125 Ala Ala Ala Lys Leu Gly Asn Ala Leu Ser Lys Met Met Gly Glu Thr 135 130

Ile Lys Lys Leu Val Pro Asn Val Leu Lys Gln Leu Ala Gln Asn Gly 150 155 Ser Lys Leu Phe Thr Gln Gly Met Gln Arg Ile Thr Ser Gly Leu Gly 165 170 Asn Val Gly Ser Lys Met Gly Leu Gln Thr Asn Ala Leu Ser Lys Glu 180 185 190 Leu Val Gly Asn Thr Leu Asn Lys Val Ala Leu Gly Met Glu Val Thr 195 200 Asn Thr Ala Ala Gln Ser Ala Gly Gly Val Ala Glu Gly Val Phe Ile 215 220 Lys Asn Ala Ser Glu Ala Leu Ala Asp Phe Met Leu Ala Arg Phe Ala 230 235 Met Asp Gln Ile Gln Gln Trp Leu Lys Gln Ser Val Glu Ile Phe Gly 250 245 Glu Asn Gln Lys Val Thr Ala Glu Leu Gln Lys Ala Met Ser Ser Ala 260 265 Val Gln Gln Asn Ala Asp Ala Ser Arg Phe Ile Leu Arg Gln Ser Arg 280 Ala Glx 290

### (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 410 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Leu Ile Ser Asn Val Gly Ile Asn Pro Ala Ala Tyr Leu Asn Asn His Ser Val Glu Asn Ser Ser Gln Thr Ala Ser Gln Ser Val Ser Ala 20 25 Lys Asp Ile Leu Asn Ser Ile Gly Ile Ser Ser Lys Val Ser Asp 40 Leu Gly Leu Ser Pro Thr Leu Ser Ala Pro Ala Pro Gly Val Leu Thr Gln Thr Pro Gly Thr Ile Thr Ser Phe Leu Lys Ala Ser Ile Gln Asn 70 75 Thr Asp Met Asn Gln Asp Leu Asn Ala Leu Ala Asn Asn Val Thr Thr 90 85 Lys Ala Asn Glu Val Val Gln Thr Gln Leu Arg Glu Gln Gln Ala Glu 100 105 Val Gly Lys Phe Phe Asp Ile Ser Gly Met Ser Ser Ala Val Ala 120 125 Leu Leu Ala Ala Ala Asn Thr Leu Met Leu Thr Leu Asn Gln Ala Asp 135 140 Ser Lys Leu Ser Gly Lys Leu Ser Leu Val Ser Phe Asp Ala Ala Lys 150 155 Thr Thr Ala Ser Ser Met Met Arg Glu Gly Met Asn Ala Leu Ser Gly 165 170 175 Ser Ile Ser Gln Ser Ala Leu Gln Leu Gly Ile Thr Gly Val Gly Ala 180 185 Lys Leu Glu Tyr Lys Gly Leu Gln Asn Glu Arg Gly Ala Leu Lys His 200 Asn Ala Ala Lys Ile Asp Lys Leu Thr Thr Glu Ser His Ser Ile Lys 215 220 210 Asn Val Leu Asn Gly Gln Asn Ser Val Lys Leu Gly Ala Glu Gly Val 230 225

Asp Ser Leu Lys Ser Leu Asn Met Lys Lys Thr Gly Thr Asp Ala Thr 250 Lys Asn Leu Asn Asp Ala Thr Leu Lys Ser Asn Ala Gly Thr Ser Ala 260 265 Thr Glu Ser Leu Gly Ile Lys Asp Ser Asn Lys Gln Ile Ser Pro Glu 275 280 285 His Gln Ala Ile Leu Ser Lys Arg Leu Glu Ser Val Glu Ser Asp Ile 295 Arg Leu Glu Gln Asn Thr Met Asp Met Thr Arg Ile Asp Ala Arg Lys 310 315 320 Met Gln Met Thr Gly Asp Leu Ile Met Lys Asn Ser Val Thr Val Gly 325 330 Gly Ile Ala Gly Ala Ser Gly Gln Tyr Ala Ala Thr Gln Glu Arg Ser 345 340 Glu Gln Gln Ile Ser Gln Val Asn Asn Arg Val Ala Ser Thr Ala Ser 360 365 Asp Glu Ala Arg Glu Ser Ser Arg Lys Ser Thr Ser Leu Ile Gln Glu 375 380 Met Leu Lys Thr Met Glu Ser Ile Asn Gln Ser Lys Ala Ser Ala Leu 390 395 Ala Ala Ile Ala Gly Asn Ile Arg Ala Glx 405

### (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 344 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Leu Asn Ile Gln Asn Tyr Ser Ala Ser Pro His Pro Gly Ile Val Ala Glu Arg Pro Gln Thr Pro Ser Ala Ser Glu His Val Glu Thr Ala 20 25 Val Val Pro Ser Thr Thr Glu His Arg Gly Thr Asp Ile Ile Ser Leu 40 Ser Gln Ala Ala Thr Lys Ile His Gln Ala Gln Gln Thr Leu Gln Ser 55 Thr Pro Pro Ile Ser Glu Glu Asn Asn Asp Glu Arg Thr Leu Ala Arg 70 75 Gln Gln Leu Thr Ser Ser Leu Asn Ala Leu Ala Lys Ser Gly Val Ser 85 90 Leu Ser Ala Glu Gln Asn Glu Asn Leu Arg Ser Ala Phe Ser Ala Pro 100 105 Thr Ser Ala Leu Phe Ser Ala Ser Pro Met Ala Gln Pro Arg Thr Thr 115 120 125 Ile Ser Asp Ala Glu Ile Trp Asp Met Val Ser Gln Asn Ile Ser Ala 135 140 Ile Gly Asp Ser Tyr Leu Gly Val Tyr Glu Asn Val Val Ala Val Tyr 150 155 Thr Asp Phe Tyr Gln Ala Phe Ser Asp Ile Leu Ser Lys Met Gly Gly 165 170 175 Trp Leu Leu Pro Gly Lys Asp Gly Asn Thr Val Lys Leu Asp Val Thr 185 Ser Leu Lys Asn Asp Leu Asn Ser Leu Val Asn Lys Tyr Asn Gln Ile 200 195 205 Asn Ser Asn Thr Val Leu Phe Pro Ala Gln Ser Gly Ser Gly Val Lys 210 215 220

Val Ala Thr Glu Ala Glu Ala Arg Gln Trp Leu Ser Glu Leu Asn Leu 230 235 Pro Asn Ser Cys Leu Lys Ser Tyr Gly Ser Gly Tyr Val Val Thr Val 245 250 255 Asp Leu Thr Pro Leu Gln Lys Met Val Gln Asp Ile Asp Gly Leu Gly 260 265 270 Ala Pro Gly Lys Asp Ser Lys Leu Glu Met Asp Asn Ala Lys Tyr Gln 275 280 285 Ala Trp Gln Ser Gly Phe Lys Ala Gln Glu Glu Asn Met Lys Thr Thr 290 295 300 Leu Gln Thr Leu Thr Gln Lys Tyr Ser Asn Ala Asn Ser Leu Tyr Asp 310 315 Asn Leu Val Lys Val Leu Ser Ser Thr Ile Ser Ser Ser Leu Glu Thr 325 330 Ala Lys Ser Phe Leu Gln Gly Glx 340

### (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 64 amino acids (B) TYPE: amino acid

  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Val Thr Ser Val Arg Thr Gln Pro Pro Val Ile Met Pro Gly Met 10 Gln Thr Glu Ile Lys Thr Gln Ala Thr Asn Leu Ala Ala Asn Leu Ser 20 25 Ala Val Arg Glu Ser Ala Thr Ala Thr Leu Ser Gly Glu Ile Lys Gly 40 Pro Gln Leu Glu Asp Phe Pro Ala Leu Ile Lys Gln Ala Ser Leu Asp 50

# (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 483 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGCATTATT	TTTTTATCAT	CGTAATCTGG	TTGCTTAGCA	TAAATACGGC	ATGGGCTGAT	60
TCTGGCTTCA	GGCTGAAAAA	ATGTTCAATA	TTGAATCCGA	ACTACTTTAC	GCTATCGCCC	120
AGCAGGAATC	GGCGATGAAA	CCTGGCGCCA	TTGGTCATAA	CCGAGATGGT	TCAACCGATC	180
TTGGCCTGAT	GCAAATTAAC	AGCTTCCATA	TGAAAAGGCT	GAAAAAAATG	GGGATTAGTG	240
AAAAACAGTT	GTTACAGGAC	CCCTGCATTT	CTGTCATTGT	GGGCGACCTC	CATTTTATCA	300
GATATGATGA	<b>AAATCTACGG</b>	TTATAGCTGG	GAGGCCGTTG	GCGCTTATAA	TGCCGGGACG	360
TCGCCGAAAC	GATCGGATAT	AAGGAAACGT	TATGCTAAAA	<b>AAATTTGGGA</b>	GAATTACAGA	420
AAATTAAAAG	GAATGTCAGC	AGAAGAGAAA	AACAAAAGAC	TTTCTATCGC	GGCAAACAAA	480
TAA						483

### (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 579 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATCAGCTTG	C CGTCGTCATA	AGCAACTGGG	CTTGCATTGC	TTTTAGTTGT	ACAAACTGTG	60
CAGGCGTCT	T CCAGCATTCT	ATTGTTCCGT	GAATCCGGAA	ATCTGCACGT	ACCTGCTCCA	120
GATTACTAT	G AGGATTATCC	TTAAGTACAA	GGGCCGCCGC	CATCGTTCCG	GTTCTTCCCA	180
CTCCGCCCA	G ACAATGAATC	ATCGGTAAAT	GCTTATCTGA	TGAACTACGC	CCCGGCGCGC	240
CATTTTGGT	I ACTATTTTC	ACCCTATCCG	CCAGGTATTC	TAACTGATCC	GTAGACGGTA	300
ACGGCTGGT	G ATCTGGCCAA	TTTTTCACAT	GCAATACCGG	GATTGTATAC	CGCTTTCCCC	360
GCAGGACAG	T TGCATATTGT	ATTGGTCTAT	CGCTTCTCCC	TGACTGGCTG	AGCTCTCTTT	420
TTGGCTGTT	G GTATGCACCT	CGCCAAAGGT	GTAGCTCCCT	CTGAAATAGG	TGGTAATTGT	480
TTTGCCTGC	A TCTGATCTTC	CGACGTTAAC	ACCACCAGGC	ACGAGCATTC	TTTTTCAAGA	540
AGCATTTTC	A TATGCGCTTC	CAGCGCATCC	CGGCGATTT			579

### (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 160 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met His Tyr Phe Phe Ile Ile Val Ile Trp Leu Leu Ser Ile Asn Thr 10 1 Ala Trp Ala Asp Ser Gly Phe Arg Leu Lys Lys Cys Ser Ile Leu Asn 20 25 Pro Asn Tyr Phe Thr Leu Ser Pro Ser Arg Asn Arg Arg Glx Asn Leu 45 35 40 Ala Pro Leu Val Ile Thr Glu Met Val Gln Pro Ile Leu Ala Glx Cys 55 Lys Leu Thr Ala Ser Ile Glx Lys Gly Glx Lys Lys Trp Gly Leu Val 70 75 Lys Asn Ser Cys Tyr Arg Thr Pro Ala Phe Leu Ser Leu Trp Ala Thr 85 90 Ser Ile Leu Ser Asp Met Met Lys Ile Tyr Gly Tyr Ser Trp Glu Ala 100 105 110 Val Gly Ala Tyr Asn Ala Gly Thr Ser Pro Lys Arg Ser Asp Ile Arg 115 120 125 Lys Arg Tyr Ala Lys Lys Ile Trp Glu Asn Tyr Arg Lys Leu Lys Gly 135 140 Met Ser Ala Glu Glu Lys Asn Lys Arg Leu Ser Ile Ala Ala Asn Lys 145 150

# (2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 239 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: peptide

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Trp Pro Gly Thr Ile Cys Gly Gln Gln His Ser Ile Asn Gln Gln Thr 10 Gln Val Lys Leu Ser Asp Gly Met Pro Val Pro Val Ile Arg Leu Thr 20 25 Phe Asp Gly Lys Pro Val Ala Leu Ala Gly Ile Arg Thr Gln Lys Ile 35 40 Arg Pro Asp Arg Leu Glu Ala His Met Lys Met Leu Leu Glu Lys Glu 55 60 Cys Ser Cys Leu Val Val Leu Thr Ser Glu Arg Ser Asp Ala Gly Lys 70 Thr Ile Thr Thr Tyr Phe Arg Gly Ser Tyr Thr Phe Gly Glu Val His 90 85 Thr Asn Ser Gln Lys Val Ser Ser Ala Ser Gln Gly Glu Ala Ile Asp 105 110 100 Gln Tyr Asn Met Gln Leu Ser Cys Gly Glu Lys Arg Tyr Thr Ile Pro 120 125 Val Leu His Val Lys Asn Trp Pro Asp His Gln Pro Leu Pro Ser Thr 135 130 140 Asp Gln Leu Glu Tyr Leu Ala Asp Arg Val Lys Asn Ser Asn Gln Asn 150 Gly Ala Pro Gly Arg Ser Ser Ser Asp Lys His Leu Pro Met Ile His 175 170 165 Cys Leu Gly Gly Val Gly Arg Thr Gly Thr Met Ala Ala Leu Val 185 180 Leu Lys Asp Asn Pro His Ser Asn Leu Glu Gln Val Arg Ala Asp Phe 195 200 205 Arg Ile His Gly Thr Ile Glu Cys Trp Lys Thr Pro Ala Gln Phe Val 215 220 210 Gln Leu Lys Ala Met Gln Ala Gln Leu Leu Met Thr Thr Ala Ser 230 235

# (2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1538 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ATGCGTGATT	GCCTGAATAA	CGGCAATCCA	GTGCTTAACG	TGGGAGCGTC	AGGTCTTACC	60
ACCTTACCAG	ACCGTTTACC	ACCGCATATT	ACAACACTGG	TTATTCCTGA	TAATAATCTG	120
ACCAGCCTGC	CGGAGTTGCC	GGAAGGACTA	CGGGAGCTGG	AGGTCTCTGG	TAACCTACAA	180
CTGACCAGCC	TGCCATCGCT	GCCGCAGGGA	CTACAGAAGC	TGTGGGCCTA	TAATAATTGG	240
CTGGCCAGCC	TGCCGACGTT	GCCGCCAGGA	CTAGGGGATC	TGGCGGTCTC	TAATAACCAG	300
CTGACCAGCC	TGCCGGAGAT	GCCGCCAGCA	CTACGGGAGC	TGAGGGTCTC	TGGTAACAAC	360
CTGACCAGCT	GCGCGCGCTG	CCGTCAGGAC	TACAGAAGCT	GTGGGCCTAT	AATAATCGGC	420
TGACCAGCCT	GCCGGAGATG	TCGCCAGGAC	TACAGGAGCT	GGATGTCTCT	CATAACCAGC	480
TGACCCGCCT	GCCGCAAAGC	CTCACGGGTC	TGTCTTCAGC	GGCACGCGTA	TATCTGGACG	540
GGAATCCACT	GTCTGTACGC	ACTCGTGACA	GGCTCTGCGG	ACATCATTGG	CCATTCAGGC	600
<b>ATCAGGATAC</b>	<b>ACTTCGATAT</b>	GGCGGGGCCT	TCCGTCCCCG	GGAAGCCCGG	GCACTGCACC	660
TGGCGGTCGC	TGACTGGCTG	ACGTCTGCAC	GGGAGGGGA	AGCGGCCCAG	GCAGACAGAT	720
GGCAGGCGTT	CGGACTGGAA	GATAACGCCG	CCGCCTTCAG	CCTGGTCCTG	GACAGACTGC	780
GTGAGACGGA	AAACTTCAAA	AAAGACGCGG	GCTTTAAGGC	ACAGATATCA	TCCTGGCTGA	840
CACAACTGGC	TGAAGATGCT	GCGCTGAGAG	CAAAAACCTT	TGCCATGGCA	ACAGAGGCAA	900

CATCAACCTG	CGAGGACCGG	GTCACACATG	CCCTGCACCA	GATGAATAAC	GTACAACTGG	960
TACATAATGC	AGAAAAAGGG	GAATACGACA	ACAATCTCCA	GGGGCTGGTT	TCCACGGGGC	1020
GTGAGATGTT	CCGCCTGGCA	ACACTGGAAC	AGATTGCCCG	GGAAAAAGCC	GGAACACTGG	1080
CTTTAGTCGA	TGACGTTGAG	GTCTATCTGG	CGTTCCAGAA	TAAGCTGAAG	GAATCACTTG	1140
AGCTGACCAG	CGTGACGTCA	GAAATGCGTT	TCTTTGACGT	TTCCGGCGTG	ACGGTTTCAG	1200
ACCTTCAGGC	TGCGGACGTT	CAGGTGAAAA	CCGCTGAAAA	CAGCGGGTTC	AGTAAATGGA	1260
TACTGCAGTG	GGGGCCGTTA	CACAGCGTGC	TGGAACGCAA	AGTGCCGGAA	CGCTTTAACG	1320
CGCTTCGTGA	AAAGCAAATA	TCGGATTATG	AAGACACGTA	CCGGAAGCTG	TATGACGAAG	1380
TGCTGAAATC	GTCCGGGCTG	GTCGACGATA	CCGATGCAGA	ACGTACTATC	GGAGTAAGTG	1440
CGATGGATAG	TGCGAAAAA	GAATTTCTGG	ATGGCCTGCG	CGCTCTTGTG	GATGAGGTGC	1500
TGGGTAGCTA	TCTGACAGCC	CGGTGGCGTC	TTAACTGA			1538

# (2) INFORMATION FOR SEQ ID NO:14

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 513 amino acids
    (B) TYPE: amino acid
    (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Met	Arg	Asp	Сув	Leu	Asn	Asn	Gly	Asn	Pro	Val	Leu	Asn	Val	Gly 15	Ala
Ser	Gly	Leu	Thr 20	Thr	Leu	Pro	Asp	Arg 25		Pro	Pro	His	Ile 30		Thr
Leu	Val	Ile 35	Pro	Asp	Asn	Asn	Leu 40	Thr	Ser	Leu	Pro	Glu 45	Leu	Pro	Glu
Gly	Leu 50	Arg	Glu	Leu	Glu	Val 55	Ser	Gly	.Asn	Leu	Gln 60	Leu	Thr	Ser	Leu
Pro 65	Ser	Leu	Pro	Gln	Gly 70	Leu	Gln	Lys	Leu	Trp 75	Ala	Tyr	Asn	Asn	Trp 80
Leu	Ala	Ser	Leu	Pro 85	Thr	Leu	Pro	Pro	Gly 90	Leu	Gly	Asp	Leu	Ala 95	Val
Ser	Asn	Asn	Gln 100	Leu	Thr	Ser	Leu	Pro 105	Glu	Met	Pro	Pro	Ala 110	Leu	Arg
		115			_		120					125	Ala		
	130					135					140		Thr		
145					150					155			His		160
		_		165					170				Ala	175	
	-		180	_				185		-			Asp 190		
-	-	195		_			200			_		205	Arg		
_	210		_		_	215					220		Ala		
225	_				230	_				235			Ala		240
_				245					250				Ser	255	
	_	_	260	_				265					Ala 270		
_		275				_	280					285	Asp		
Leu	Arg 290	Ala	Lys	Thr	Phe	Ala 295	Met	Ala	Thr	Glu	Ala 300	Thr	Ser	Thr	Сув

Glu Asp Arg Val Thr His Ala Leu His Gln Met Asn Asn Val Gln Leu 315 310 Val His Asn Ala Glu Lys Gly Glu Tyr Asp Asn Asn Leu Gln Gly Leu 325 330 335 Val Ser Thr Gly Arg Glu Met Phe Arg Leu Ala Thr Leu Glu Gln Ile 345 340 350 Ala Arg Glu Lys Ala Gly Thr Leu Ala Leu Val Asp Asp Val Glu Val 365 355 360 Tyr Leu Ala Phe Gln Asn Lys Leu Lys Glu Ser Leu Glu Leu Thr Ser 375 380 Val Thr Ser Glu Met Arg Phe Phe Asp Val Ser Gly Val Thr Val Ser 390 395 Asp Leu Gln Ala Ala Asp Val Gln Val Lys Thr Ala Glu Asn Ser Gly 405 410 Phe Ser Lys Trp Ile Leu Gln Trp Gly Pro Leu His Ser Val Leu Glu 420 425 Arg Lys Val Pro Glu Arg Phe Asn Ala Leu Arg Glu Lys Gln Ile Ser 440 445 Asp Tyr Glu Asp Thr Tyr Arg Lys Leu Tyr Asp Glu Val Leu Lys Ser 450 455 460 Ser Gly Leu Val Asp Asp Thr Asp Ala Glu Arg Thr Ile Gly Val Ser 475 470 Ala Met Asp Ser Ala Lys Lys Glu Phe Leu Asp Gly Leu Arg Ala Leu 490 485 Val Asp Glu Val Leu Gly Ser Tyr Leu Thr Ala Arg Trp Arg Leu Asn 500 505 Glx

### (2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3441 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CGCAAAGCCG	AGGAAACGAA	CCGCATTATG	<b>GGATGTATCG</b>	GGAAAGTCCT	CGGCGCGCTG	60
CTAACCATTG	TCAGCGTTGT	GGCCGCTGTT	TTTACCGGTG	GGGCGAGTCT	GGCGCTGGCT	120
GCGGTGGGAC	TTGCGGTAAT	GGTGGCCGAT	GAAATTGTGA	AGGCGGCGAC	GGGAGTGTCG	180
TTTATTCAGC	AGGCGCTAAA	CCCGATTATG	GAGCATGTGC	TGAAGCCGTT	AATGGAGCTG	240
ATTGGCAAGG	CGATTACCAA	AGCGCTGGAA	GGATTAGGCG	TCGATAAGAA	AACGGCAGAG	300
ATGGCCGGCA	GCATTGTTGG	TGCGATTGTC	GCCGCTATTG	CCATGGTGGC	GGTCATTGTG	360
GTGGTCGCAG	TTGTCGGGAA	AGGCGCGGCG	GCGAAACTGG	GTAACGCGCT	GAGCAAAATG	420
ATGGGCGAAA	CGATTAAGAA	GTTGGTGCCT	<b>AACGTGCTGA</b>	<b>AACAGTTGGC</b>	GCAAAACGGC	480
AGCAAACTCT	TTACCCAGGG	GATGCAACGT	ATTACTAGCG	GTCTGGGTAA	TGTGGGTAGC	540
AAGATGGGCC	TGCAAACGAA	TGCCTTAAGT	AAAGAGCTGG	TAGGTAATAC	CCTAAATAAA	600
GTGGCGTTGG	GCATGGAAGT	CACGAATACC	GCAGCCCAGT	CAGCCGGTGG	TGTTGCCGAG	660
GGCGTATTTA	TTAAAAATGC	CAGCGAGGCG	CTTGCTGATT	TTATGCTCGC	CCGTTTTGCC	720
ATGGATCAGA	TTCAGCAGTG	GCTTAAACAA	TCCGTAGAAA	TATTTGGTGA	AAACCAGAAG	780
GTAACGGCGG	<b>AACTGCAAAA</b>	AGCCATGTCT	TCTGCGGTAC	AGCAAAATGC	GGATGCTTCG	840
CGTTTTATTC	TGCGCCAGAG	TCGCGCATAA	AAACTGCCAA	<b>AATAAAGGGA</b>	GAAAAATATG	900
TTAATTAGTA	ATGTGGGAAT	AAATCCCGCC	<b>GCTTATTTAA</b>	ATAATCATTC	TGTTGAGAAT	960
<b>AGTTCACAGA</b>	CAGCTTCGCA	ATCCGTTAGC	GCTAAAGATA	TTCTGAATAG	TATTGGTATT	1020
AGCAGCAGTA	<b>AAGTCAGTGA</b>	CCTGGGGTTG	AGTCCTACAC	TGAGCGCGCC	TGCGCCAGGG	1080
GTATTAACGC	AAACCCCCGG	AACGATCACG	TCCTTTTTAA	AAGCCAGTAT	TCAAAATACC	1140
GACATGAATC	AGGATTTGAA	TGCTCTGGCA	<b>AATAATGTCA</b>	CGACTAAAGC	GAATGAGGTT	1200
GTGCAAACCC	<b>AGTTACGCGA</b>	GCAGCAGGCA	GAAGTCGGAA	AGTTTTTTGA	TATTAGCGGA	1260

CAGGCTGATA CCATGATCC CCATGATCCC CGAAGCGT CCCATAAACA TAATGCCGCC CGCAACTGCA AATGCATCA CGCTTCAAACA TAATGCCGCC AAGATCGATA AACTGACCA CGTTAAAAA CCGTCTCAA CGGCCACGCTA AACTCAAAAAA CCGTCTCAAA CCGTCAAAAC CGGCACCCTA AATCTAAATCA CGGCACCCCTA AATCTAAATCA CGGCACCCCTA AATCTAAATCA CGGCACCCCTA AATCTAAATCA CGGCACCCCTA CCCCACCGAAA ACCCCACCGAAA ACCCCCCCCACC CCCACCGAAA ACCCCCCCC	<b>አ</b> ጥርጥርጥጥርር እ	GTGCCGTTGC	GCTGTTGGCT	GCCGCGAATA	CGTTAATGCT	GACGTTGAAC	1320
ACGGCAAGGT CCATGATGC CGAAGGGATG AATGCGTTGT CCGGTAGTAT TTCCCAGAGC GCACAGGT TGGGGATCAC TGGCGTGGGG GCCAAACTGG AATATAAGGG GCTGCAGAAT 1500 GAAAGAGGC CGCTAAACA TAATGCGCGG AGATGGATA AACTGACCAC TGAAAGCAC 1560 AGTATTAAAA ACGTCTGAA CAGGCAGAAT AGCGCACACAC TCGAAAAA TCTTAATGAT GAACAAAACC GGGCACACATA AACTCAACACAC TCGGTGATA TCTTAATGAT 1620 TCTCTGAAAT CGTTAAATAT GAACAAAACC GGTACCGATG CGACAGAAAA TCTTAATGAT 1680 GCGACGCTTA AACTCAACAC CGGAACACAC CCACGGAAA GTCTGGGTAT TAAACACAGT 1740 AATAAACAAA TCTCCCCTGA ACATCAGGCT ATTCTGTCGA AACGTCTTGA GTCTTCGAA 1800 TCCCATATTC GTCTTAGAGA GAATACCATG GATATGACCC GAATCGATG GCGCAAGATG 1800 TCCCATATTC GTCTTAGACA TATGAAGAC TCGGTACGGT TCGGGTAT TGAACACAGT 1800 TCCCGGGACG GCACCGCAAAA TCTCCCCTGA ACATCAGGC CGAACACCC GAATCGATG GCGCAAGATG 1800 TCCCGGGCAG ACCCGCATA TATGAAGAC TCCGGTCACGG TCGGTGGTAT TGCAGGGGC 1920 TCCGGGCAGAT ACGCCGCTAC GGACCAACACC CGCACACACC GAATCGATG GCGCAACACC CGCACACAC CGCACACACC GCACCGCACA CACCGCACAC GCACCGCACAC GCACCACACC GGACCAACAC AATGCAGACC CTGAAAACA TCCACAGAC AATGCACAC CAAAACAATC CACCACACAC ACTCCCTCG GAACACAC ACTCCCTCG GCAATATTC CGCCTCTCTCCT CATCCGGGGA TCGTTGCCA ACGCCCCCAC CACTCCACCAC ACCCCCCCCCC						<b>4</b>	
GCGCTTCAGT TGGGGATCAC GAAAGAGGCG CGCTTAAACA TAATGCCGCG AAGATCTGC AATATAAAGG GCTGCAGAAT GAAAGAGGCG CGCTTAAACA TAATGCCGCG AAGATCTGCATA AACTGACCAC TGAAAGCCAC TGAAAGCCAC TCTCTGAAAT ACGTGCTGAA CGGGCAGAAT AGCGTCAAAC TCGGTGCTGA AGCGCTGAT TCTCTGAAAT CGTTAAATAT GAACAAAACC GGTACCGAT CGACGAAAAA TCTTAATGAT GCGACGCTTA AATCTAATGC CGGAACCAGC GCCACGGAAA GTCTGGGTAT TAAAGACAGT TCCCCATATTC GTCTTGAGCA GAATACCATC GATTATCCC GAATGATCT TAAAACAAA TCTCCCCTGA ACATCAGGCT ATTCTGTCCA AACGTCTGA GCCCACAGAAA TCTCTGCGAA TCCCGATATTC GTCTTGAGCA GAATACCATC GATTATCACC GAATGATCC GCCCAAGATG TCCGGGCAGT ACCCCGCTAC TCAGGAACGT TCCGACGAC CGACTGATAT TGCAGGGGCG 1920 TCCGGGCAGT ACCCCGCTAC TCAGGAACGT TCCGACGAC CGATTAGACCA GGTGAATAAC AATTCAGGAAA TGCTGAAAAC AATGGAGAC CGTGAAAATT TCCACGAGCAC GGTGAAAATATCC GCCCTAAAAC AATGGAGACC CGTGAAAATT CACCAGAACC CGAATAATCC GCCCTAAAAC AATGGAGACC CGTGAAAATTAGCC GCACCAGCCT 2040 ATTCAGGAAA TGCTGCAAAAC AATGGAGACC CGTGAAAATTATC CACTTAAATCC CGCCTTAATCT GAAAGATCAT CTATAACGCA TCATGGGTGT 2160 GCTATTAAACC GCCCCTGAT GGCCAAACTG GGATATATTC CACTTAAATCC CGCCTCTGAT GGCCAACTCGG GGAACATATTCC CACTCCTGAT AAAATTATTC CACTCCAGACC CGTTCACCA AGACATCCC GGTACAGACA CTTAACACCA CGCACCGCT 2280 CGTCGAGACT GCCGTGGTAC CGTTTACCAC AGAACATCCC GGTACACAAC CGCACCCGAT 2280 CGTCGAGACG GCTACTAAAA TCCACCAGC ACAGCAGACG CTGCACTACA CGCCACCGCT 2400 CTCTGAAAGA AATAATGAC AGCCCACGC GGCGCCCAG CAGTTGACCA GCCACCGGT 2400 CTCTGAAAGA AATAATGAC AGCCCACGCT GCCACCACGC CAGGTGACAC CGCACCAGC CAGCTCAGACA CACCAGACAC CTGCACTACA CGCCACCGAT 2400 CTCTGAAAAT ACCTCACAAATGG GAGCCTATAAA ACGCCACGCT TTCCCCAAAATA ATATCAGACAC CCACAGCC CAGCTCGC CACGCTCAC AGCCACACC CAGCCACCGC CAGCTCACACC CACCAGCC CAGCCCACCCACCCACCCACCCAC							
GARAGAGGCG CGCTTARACA TANTGCCGCG ARGATCGATA ACTGACCAC TGARAGCAC 1560 AGTATTARAAA ACGTGCTGAA CGGGCAGAAT AGCGTCAAAC TCGGTGGTGA AGCGGTCGAT 1620 TCTCTGAAAT CGTTARATAT GAAGAAACC GGTACCGATG CGACGAAAA TCTTAATGAT 1680 GCGACGCTTA AATCTAATGC CGGAACCAC GCACCGGAAA GTCTGAGTAT TAAAGACACT 1740 AATAAACAAA TCTCCCCTGA ACATCAGGCT ATTCTCTCGA AACGTCTTGA GTCTGTCGAA 1860 TCCGATATTC GTCTTGAGCA GAATACCATG GAATACCATG GAATCGATG GCGCAAGATG 1860 CCGAGTGCCG GCGCATCTGAT TATGAAGAC TCCGGTCACGG AAATCCATG GCGCAAGATG 1860 CCGGGTGCCA GCACCGCATC TCAGGAACC CGGTAAAACCATG GAATCCACG CGCACAGAATG 1860 CGGGTTGCCA GCACCGCATC TCAGGAACCT TCCGAGCACG AAATTAGCCA GCGCAAGATG 1920 TCCGGGCAGT ACGCCCATC TCAGGAACCT TCCGAGCACG AAATTAGCCA GCGCAAGATG 1980 CGGGTTGCCA GCACCGCATC CGACCGAAGCC CGTGAAAAGT CGCCAAGACAC CGCACTACCT TCAGGACCT CCGCATCACCT 1980 GCTATCGCAG GCAATATTCG CGCTTAATCT GAAAGGTCAT CAGTAAATC GACCAGCCTG 2040 ATTCAGGAAA TGCTGAAAA AATGGAGAC ATTAACCAGT CGAAAGCATC CGCACTCGCT 2100 GCTATCGCAG GCAATATTCG CGCTTAATCT GAAAGGTCAT CTATACCACA TCATGGGTTT 2160 GATTTAATCG CGCTCCTGAT CGCGCAACACA AGACCATCCC CGAACCACC TCATGGGTT 2160 GATTTAATACC CGCCTCGAT CGTTGCCAC AGACCATCCC CGGTAAAAC ATTAACCAC AGACCATCCC CGGTACACAA AAATAATCAC CGCCCTCGACC CGTTCACCAC AGACCATCCC CGGTACACAA AATAATCACA CCGTCCACCAC AGACCATCCC CGGTTCACCA AGACCATCCC CGGTTGACCA AATAATCACA CCGTCCACCAC CGCCCCAC CAGTTGACCA CAGCCCTGAAAA TCCACCCACCCACCACCAC CAGTTCACCA CAGCCCCACACAAA AATGAGAAC CGCCACCCGAT 2400 CTCTCAAAAA GCCGCACGCT GGCCCCCAC CAGTTGACCA CGCACCCCGAT 2400 CTCTCAAAAA AATAATCACA CCGCCACCACCAC CAGTTGACCA CAGCCCCCAAAAA AATGAGAAC CCGCACCCCACACAC AATAACCACAC CCAACAAA AATGAGAAC CCAACAAA AATGAGAAC CCGCACCCCACACAC CAGCACCAC CAGCACCAC CAGACAAA AATGAGAAC CCGCACCCCAC CAGACAAA AATGAGAAC CCGAGACAAC CCAACAAAAAAAAAA			• • • • • • • • • • • • • • • • • • • •				1500
AGTATTAAAA ACGTGCTGAA CGGGCAGAAT AGCGTCAAAC TCGGTGCTGA AGCGTCGAT 1620 TCTCTGAAAT CGTTAAATAT CAAGAAAACC GGTACCGATG CGACGAAAAA TCTTAATGAT 1680 GCGACGCTTA AATCTAATGC CGGAACCAGC GCCACGGAAA GTCTGGGTAT TAAAGACAGT 1740 AATAAACAAA TCTCCCCTGA ACATCAGGCT ATTCTGTCGA ACGTCTTGA GTCTGTGCAA 1800 TCCGGATATTC GTCTTGAGCA GAATACCATG GATATACACC GAATCGATGC GCCCAGAGAG 1860 CAGATGACGG GCGATCTGAT TATGAAGAAC TCGGTCACGA CGACGAGAG GCGCAAGATG 1860 CAGATGACGG GCGATCTGAT TATGAAGAAC TCGGTCACGA TCGGTGCTAT TGCAGGGGCG 1920 TCCCGGGCAGT ACGCCCGCTAC TCAGGAACCGT TCCGAGCAGC AAATTACCCA GGTGAATAAC 1980 CGGGTTGCCA GCACCGCATC GGACGAAGCC CGTGAAAGATC CGCACCAGC AAATTACCCA GGTGAATAAC 1980 CGGTATCGCAG GCAATATTCG CGCTCAAATCT GAAAGGTCAT CTATACCGCA TCATGGGTGT 2160 GATTTAATCG CGCTCTGAT GGCGAACCAC GGTAAAACC TCATGGGTGT 2160 GATTTAATCG CGCTCTGAT GGCGAACTG GGATATTTCG CGCACCAGCT CAAAGGTCAC TCATGGGTGT 2160 GATTTAATCG CGCTCTGAT GGCGAACTG GGATATTATG CTTAATATTC AAAATTATTC 2220 CGCTTCCCT CATCCGGGA TCGTCCCCA ACGGCCGCAG ACTCCCTGG CGAGCAGCA CCGTCGAGCAGC CGTCACACACA AAATTACCCCA GCACCACCAC ACGCCCAGCAC CCGTCCACAC ACGCCAGCAC CCGCCACCACAC ACGCCAGCAC CCGACCACCAC ACGCCAGCAC CCGACCACCAC ACGCCACCAC CCGACCACCAC ACGCCACCAC CCGACCACCAC ACGCCAGCAC CCGACCACCAC ACGCCACCAC CCGACCACCAC ACGCCACCAC CCGACCACCAC ACGCCACCAC ACGCCACCAC ACGCCACCAC ACGCCACCAC CCGACCACCAC ACGCCACCAC CCGACCACCAC ACGCCACCACAC AAACAAAAAAAA							
TCTCTGAAAT CGTTAAATAT GAAGAAAAC GGTACCGATG CGACGAAAAA TCTTAATGAT 1680 GCGACGCTTA AATCTAATGC CGGAACCAGC GCCACGGAAA GTCTGGGTAT TAAAGACAGT 1740 AATAAACAAA TCTCCCCTGA ACATCAGGCT ATTCTGTCGA ACGTCTTGA GTCTTGCAAA 1800 TCCGATATTC GTCTTGAGCA GAATACCATG GATATCACC GAATCGATGC GCGCAACATG 1860 CAGATGACGG GCGCACCT TAAGAACAC TCCGGTCACGG TCGGTGGTAT TGCAGGGGCG 1920 TCCGGGCAGT ACGCCGCTAC TCAGGAACGT TCCGACCAGC AAATTAGCCA GGTGAATAAC 1980 CGGGTTGCCA GCACCGCATC GGACGAGC CGTGAAAGTT CACGTAAATC GACCAGCCTG 2040 ATTCAGGAAA TGCTGAAAAC AATGGAGAC ATTAACCAGT CGAAAGCAT CGCACCTGCT 2100 GCTATCGCAG GCACCTCTGAT GGCGAACTG GAAAGTT CACATCAGCAC ACGCCTCCTCTC CACCGCTCCT GAAAGCTC CGCACTCGCT 2100 GCTATCCCC GCCTCCTGAT GGCGAACTG GAATATATG CTTAATATTC AAAATTATTC 2220 CGCTTCTCCT CACCGGGGA TCGTTGCCCA ACGCCCGAC ACTCCCTCGG CGAGCGACA CCCCCTCCTCCT CACCGCGGA TCGTTGCCCA ACGCCCGCA CACCCCAGC CACGCCAGC ACGCCCAGC ACGCCCAGC ACGCCCAGC ACGCCCAGC ACGCCCAGC ACGCCCAGC ACGCCCAGC ACGCCCAGC CCCACTCGC CGAGCGACA AATAATGACC ACGCCCGCA ACGCCCAGA CCCCCCAT CACCCAGGC CACGCCAGACA AATAATGACC ACGCCCACCA CACCCAGACA CACCCAGGC CACGCCAGACA AATAATGACC ACGCCACCCA CACCCAGACA CACCCAGACA CACCCAGCC CACCCAAT CACCCAGCC CACTCTCACCA AAATAATGACC ACGCCCACCCAC CACCCAGCC CACTCGACC CACCCCAT CACCCAGCC CACCCAGACAC CACCCAGACA CACCCAGACA CACCCAGCC CACCCAGACAC CACCCAGACA CACCCAGCC CACCCAGACAC CACCCAGACACA CACCCAGACAC CACCCACACAC AAACCACAT AAACCACAT AAACCACAT TACCCCACAC TACCCACAC TACCCACAC TACCCACAC CACACCACAC CACACCACAC CACACCACAC CACACCAC		•••					
GCGACGCTTA AATCTAATGC CGGAACCAGC GCCACGGAAA GTCTGGGTAT TAAAGACAGT 1740 AATAAACAAA TCTCCCCTGA ACATCAGGCT ATTCTGTCGA AACGTCTTGA GTCTGTCGAA 1860 TCCGATATTC GTCTTGAGCA GAATACCATG GATCAGACC GAATCGATGC GCCACAGATG GCCAGATGACGG GCGATCTGAT TATGAAGAAC TCGGTCACGG GAATCGATGT TGCAGGGGGG 1920 TCCGGGCAGT ACGCCGCTAC TCAGGAACGT TCCGAGCAGC TCGGTGGTAT TGCAGGGGGG 1920 CGGGTTGCCA GCACCGCATC GGACGAAGCC CGTGAAAGTT CACGTAAAC AATGGAGGAC ATTAACCAGT CGAAAGCATC GCACAGCCTG GCAATCTGCAG GCAATATTCG GCATTCAGGAAA AATGGAGGAC ATTAACCAGT CGAAAGCATC GCACTCGCT 2100 GCTATCGCAG GCAATATTCG GCCGAACTG GGAAAGTCAT CTATACGCCA TCATGGGTGT 2160 GATTTAATCG CGCTCCTGAT GGCGAACTG GGAATTATTG CTATATATTC AAAATTATTC 2220 CGCTTCCTC CATCCGGGA TCGTTGCCCA ACGCCGCGA GCTCCCCCGC GGAGCGAGCA 2280 CGTCGAGACT GCCGTGTAC ACACCAGCC ACGCCAGCAG ACTCCCTCGG GGAGCGACCA 2280 CGTCCAGAGCT GCCGTGTACAAAAA TCCACCAGCC ACACACACGC GGTACAGATA TCATTTCATT							
AATAAACAAA TCTCCCCTGA ACATCAGGCT ATTCTGTCGA AACGTCTTGA GTCTGTCGAA 1800 TCCGGATATTC GTCTTGAGCA GAATACCATG GATATGACCC GAATCCATGC GCGCAAGATG CAGATGACGG GCGATCTGAT TATGAAGAAC TCGGTCACGG TCGGTGTAT TGCAGGGGGG CAGATGACGG ACCGCCATC GCACCGAACCC TCCGGAGCACC AAATTAGCCA GGGCAACAACCC CGGGTGTACCA GCACCGCATC GGACCGAACCC CGTGAAAGCT CACCGTAAAAC ACCGCCATC GGACCGAACCC CGTGAAAGCT CACCGAAACAC CGCAATATTCG CGCTAATACCAGT CGAAACACCACCACCACCACCACCACCACCACCACCACCA		~ · · · · · · · · · · · · · · · · · · ·	•			• • • • • • • • • • • • • • • • • • • •	1740
TCCGATATTC GTCTTGAGCA GAATACCATG GATATGACCC GAATCGATGC GCGCAAGATG 1860 CAGATGACGG GCGATCTGAT TATGAAGAAC TCCGGTCACGG TCGGTGATAT TGCAGGGGGG 1920 CCCGGGCAGT ACGCCGCTAC TCAGGAACGT TCCGAGCAGC AAATTAGCCA GGTGAATAAC 1980 CGGGTTGCCA GCACCGCATC GGACCAGC CGTGAAAAGTT CACGTAAAAC GACCAGCCTG 2040 ATTCAGGAAAA TGCTGAAAAAC AATGGAGAGC CGTGAAAAGTT CCACGTAAAAC GACCAGCCTG 2100 GCTATCGCAG GCACCTGAT GGCGAACTG GGAAAGCATC CGCACTCGCT 2100 GCTATTAATCG CGCTCCTGAT GGCGAACTG GGAAAGCATC CTATACGCCA TCATGGGTGT 2160 GATTTAATCG CGCTCCTGAT GGCGAACTGG GGAAATATTC CACCAGGA TCGTTAATATTC CACCAGGCA TCGTTAAAAA TCCACCAGGC ACGCCCCAGA ACTCCCTCGG CGGCGCGCAG ACTCCCTCGG CGGCGCAGA CCGCCCAGACAC ACGCCCAGACAC CGGCGCAGA ACGCCAGACAC CGGCGCAGA AATGAGAACAC CGCCACCGAT 2400 CTCTGAAGAG AATAATCACG GGCGACGCC GGCGCCCAG CAGGTCAAAA TCATTTCATC CGCACACACA AATGAGAACA CGCCACGAT 2400 CTCTGAAGAG AATAATCACG GCGACTGAG CCGCACCAGAA AATGAGAACA CGCCACCGAT 2400 CTCTGAAGAG AATAATCACG CCTATATAC CGCACACAAA AATGAGAACC TGCGCAGCCGC TGCATTATC CGCACACAAA AATGAGAACC TGCGAGCAGAC CAGCTGAAA AATGAGAACC TGCGAGCAGAC CAGCTGAAA AATGAGAACC TGCGAGCAGAC CAGCTGAAA AATGAGAACC TGCGAGACAA CAGCCAGAACAA CAGCCAGACAA CAGCCAGAAAA AATGAGAACC TGCGAGACAA CAGCCAGAACAA CAGCCAGAACAA CAGCCAGAACAA CAGCCAGAAAA AATGAGAACC TGCGAGACAA CAGCCAGAACAA CAGCCAGACAA CAGCCAGAACAA CAGCCAGAACAA CAGCCAGAACAA CAGCCAGAACAA CAGCCAGAACAA CAGCCAGAACAA CAGCCAGAACAA CAGCCAGAAAAA ATAACCAGATAAAAA ATAACCAGATAAAAAAAAAA			•	ATTCTGTCGA	AACGTCTTGA	GTCTGTCGAA	1800
CAGATGACGG GCGATCTGAT TATGAAGAAC TCGGTCACGG TCGGTGGTAT TGCAGGGGCG 1920 TCCGGGCAGT ACGCCGCTAC TCAGGAACGT TCCGAGCAGC AAATTAGCCA GGTGAATAAC 1980 CGGGTTGCCA GCACCGCATC GGACGAGCC CGTGAAAGTT CACGTAAATC GACCAGCCTG 2040 ATTCAGGAAA TGCTGAAAAC AATGGAAGCC CGTGAAAGTT CACGTAAATC GCCACTCGCT 2100 GCTATCGCAG GCAATATTCG CGCTTAATCT GAAAGGTCAT CTATACCGCA TCATGGGTGT 2160 GATTTAATCG CGCTCCTGAT GGCGAACTGG GGATATTATG CTTAATATTC AAAATTATTC 2220 CGCTCGCTCTCCT CATCCGGGGA TCGTTGCCGA ACGCCCGCAG ACTCCCTCGG CGAGCGAGCA 2280 CGTCGAGACT GCCGTGGTAC CGCTCTACCAC ACAGCAGCG CGGACAGATA TCATTTCATT		• •		GATATGACCC	GAATCGATGC	GCGCAAGATG	1860
TCCGGGCAGT TCCGGGCAGT TCCGGGCAGC TCCGGGTTGCCA GCACCGCATC GGACGAAGCC TCGGAAAGC ATTACCGAG ATTACCAGA TCCTGAAAAC AATGAGAGAC GCTATTCC GCTTATTCC GCTTATCC GCTTCCT CATCCGGGA TCGTTGCCGA ACCGCCAC GCTTCTCCT CATCCGGGA TCGTTGCCGA ACGCCCCCAC CGGCACCCC CGCTCCTCGC CGCTCCTCC CGCTCTCCC CGCTCTCCC CGCTCCTCC CGCTCCCCA CGCCCCCCC CCTCCCCC CCGCCCCCC CCGCCCCCC CCGCCCCCC CCGCCCCCC				TCGGTCACGG	TCGGTGGTAT	TGCAGGGGCG	1920
CGGGTTGCCA GCACCGCATC ATTCAGGAAA TGCTGAAAAC AATGGAGAGC CGTTAATCT GCAAAGCCCT GCAAAGCCCC CGCACTCGCT CGCACTCGCT CGCATTCCCG GCAATATTCG CGCTCCTGAT CGCTCCTGAT CGCTCCTGAT CGCTCCTGAT CGCTCCTGCT CATCCGGGA CGCTCGCAC CGCTCCTGCT CGCTCCTCCT CATCCGGGA CGCTCGCAC CGCTCTCCC CATCCCGGGA CGCTCGCAC CGCTCGCAC CGCTCCCCC CGCTCGCCC CGCTCGCCC CGCTCGCCC CGCTCGCCC CGCTCGCCC CGCTCGCCC CGCTCGCCC CGCTCGCCC CGCCCCCCC CGCTCGCCC CGCTCGCCC CGCTCGCCC CGCTCGCCC CGCTCGCCC CGCTCGCC CGCTCGCCC CGCCCCCCC CGCCCCCCC CCCCCCCC	••••	ACGCCGCTAC	TCAGGAACGT	TCCGAGCAGC	AAATTAGCCA	GGTGAATAAC	1980
ATTCAGGAAA TGCTGAAAAC GCATATTCG GCATATCG GCATATTCG CGCTCCTGAT GGCGAACTGG GGATATTTCG CGCTCCTGAT GGCGAACTGG GGATATTATC CGCTCCCCC CATCCGGGGA CCGTCGACC CCGCCGCGCCCCC CGCCCCCGC CCGCCCCCCC CCGCCG			GGACGAAGCC	CGTGAAAGTT	CACGTAAATC	GACCAGCCTG	2040
GCTATCGCAG GCAATATTCG CGCTTAATCT GAAAGGTCAT CTATACGCCA TCATGGGTGT 2200 CGCTCTCTCT CATCCGGGA TCGTTGCCGA ACGGCCGCAG ACTCCCTCGG CGAGCAGCA 2280 CGTCGAGACT GCCGTGTACCAC AGAACATCGC GGTACAGATA TCATTCATT 2340 ATCGCAGGCG GCTACTAAAA TCCACCAGGC ACAGCAGCG CTGCAGTCAA CGCCACCGAT 2400 CTCTGAAGAG AATAATGACG AGCGCACGCT GGCGCGCCGC CAGTTGACCA GCCACCGAT 2400 CTCTGAAGAG AATAATGACG AGCGCACGCT GGCGCGCCGC CAGTTGACCA GCCACCGAT 2400 CTCTGAAGAG AATAATGACG AGCGCACGCT GGCGCCCCGA CAGTTGACCA GCCACCGAT 2400 CTCTGAGAGA AATAATGACG AGCGCACGCT GGCGCCCCGA CAGTTGACCA GCCACCGAT 2400 CTCTTGAGAGA AATAATGACG CCTTATTTC CGCAGAACAA AATGAGAACC TGCGGAGCGC 2520 GTTTTCTGAT GCTGAGATTT GGGATATGT TTCCCAAAAAT ATATCGGCGA TAGGTGACA 2460 CTATCTGGG GTTTATGAAA ACGTTGTCGC AGTCTATACC GATTTTTATC AGGCCTTCAGCT ATACCGCGA AGGCCTGAA AACGTTATA ACCGTTATAT ACCGTTAAT ACCGTTAAT AAACAGTTTA ACCGTTAAT ACCGTTAAT ACCGTTAAT AAACAGTTAA ACCGTTTAT TTCCAAGCGCA GTCAGCAGC GGCGTTAAAG AAAACAGTAAT ACCGTTTAT TTCCAAGCGCA GTCAGCAGC GGCGTTAAAA ATAATCAAAT 2820 AGGCGAAGCA AGACAGTGGC TCAGTGAATT GAATTACCG AATAACCACTAA ATAATCAAAT 2820 AGCGGAAGCA AGACAGTGGC TCAGTGAATT GAACTCCAA ATACCGTTAA ATAATCAAAT 2820 AGCGGAAGCA AGACAGTGGC TCAGTGAATT GAACTCCAA ATAACCACAT TAGCCACTGA AGCCCACGA AGACAGTGC TCAGTGAATT GAACAGTATA ATAATCAAAT 2940 AGCGCAGAAAA ATAACAATTA ACGGTTAAA AGCGCAGAA AAAACACCACA TTCAGGACAC CCAAATATCA 3000 AGCGCAAAAAA TAAGCAACG CCAATATCAT GAACACCACAT TACAGACCCT TACAGACCAC TACAGACAC CCAAATATCA AGCACCACAAAAAAAAAA			AATGGAGAGC	ATTAACCAGT	CGAAAGCATC	CGCACTCGCT	2100
GATTTAATCG CGCTCCTGAT GGCGAACTGG GGATATTATG CTTAATATTC AAAATTATTC 2220 CGCTTCTCCT CATCCGGGGA TCGTTGCCGA ACGGCCGCAG ACTCCCTCGG CGAGCGAGCA 2280 CGTCGAGACT GCCGTGGTAC CGTCTACCAC AGAACATCGC GGTACAAAAA TCCACCAGGC ACAGCAGACG CTGCAGTCAA CGCCACCGAT 2400 CTCTGAAGAG AATAATGACG AGGCCACGCT GGCGCCCAG CAGTTGACCA GCAGCCGAT 2400 CGTTTCTCGCG CCGACGCGC TGTCAATATC CGCCACCAGC CAGTTGACCA GCAGCCTGAA AATGAGAACC TGCGGAGCGC CCTTATTTAC CGCCACCAGA AATGAGAACC TGCGGAGCGC CTTATTTAC CGCTTCGCCT ATGCGCAGCC CGAGAACAAC 2580 CATTCTGAT GCTGAGATTT GGGGATATGGT TCCCAAAAT ATATCGGCGA TAGGTGACAG 2500 CTTATTTATC ACCTCTCA AAAATGATTT AAACCAGTTA ACCTCTCACAAAT ACCTCTCA AAAATGATTT AAACCAGTTA ACCTCTCA AAAATGATTT AAACCAGTTA ACCTCTCA AAAATGATTT ACCAGCGA GGCGTTAAAA TAATCCGTTAA ACCGTTAAAA ACCGTTATA CCCGTTGACC GAGAACAAC AATGAGAACC TGCGGAGCC CGAGAACAAC 2580 CTTATTTATC AGGCCTTCAG 2700 AAACCAGTATA ACCGTTTAT TCCAAAATGG GAGGCTGGTT ATTACCAGGT AAACCAGTTAA ACCGTTAAA ACCGTTTAA ACCGTTAAA ACCGTTTAA ACCGTTAAAA ACCGTTTAT TACCAGCGA GGCGTTAAAAT ATAACCAATA ATAACCAATA ATAACCAATA ATAACCAATA AAACAGTTAA ACCGCTAAA AAACAGTTAA CCCGCTAAAATA ACCGCCAAAAAAAATGG TCCAAAAAAAATGG TTCAAGAATCT AAAACCACAT TACAGACCGC AAAACCACAT TACAGACCGC AAAACCACAT TACAGACCGC AAAACCACAT TACAGACCGC AAACCGCCAA AACCGCCAA AACCGCCAA AACCGCCAA AACCGCCAA AACCACAT TACAGACCGC AAAACCACAT TACAGACCGC AAAACCACAT TACAGACCGC AAACCACCAT AAAACCACAT TACAGACCGC AAACCGCCAA AACCGCCAA AACCGCCAC CCGAGAACAAC CCGAAAAAAAA				GAAAGGTCAT	CTATACGCCA	TCATGGGTGT	2160
CGCTTCTCCT CATCCGGGA TCGTTGCCGA ACGCCCGCAG ACTCCCTCGG CGAGCGAGCA 2280 CGTCGAGACT GCCGTGGTAC CGTCTACCAC AGAACATCGC GGTACAGATA TCATTTCATT		CGCTCCTGAT	GGCGAACTGG	GGATATTATG	CTTAATATTC	AAAATTATTC	2220
CGTCGAGACT GCCGTGGTAC CGTCTACCAC AGAACATCGC GGTACAGATA TCATTCATT 2340 ATCGCAGGCG GCTACTAAAA TCCACCAGGC ACAGCAGACG CTGCAGTCAA CGCCACCGAT 2400 CTCTGAAGAG AATAATGACG AGCGCACGCT GGCGCGCCAG CAGTTGACCA GCAGCCTGAA 2460 TGCGCTGGCG AAGTCCGGCG TGTCATTATC CGCAGAACAA AATGAGAACC TGCGGAGCGC 2520 GTTTTCTGCG CCGACGTCGG CCTTATTTAG CGCTTCGCCT ATGCGCAGC CGAGAACAAC 2580 CTATCTGGT GCTGAGATTT GGGATATGGT TTCCCAAAAT ATATCGGCGA TAGGTGACAG 2640 CTATCTGGGC GTTTATGAAA ACGTTGTCGC AGTCTATACC GATTTTATC AGGCCTTCAG 2700 TGATATTCTT TCCAAAATGG GAGGCTGGTT ATTACCAGGT AAGGACGGTA ATACCGTTAA 2760 GCTAGATGTT ACCTCACTCA AAAATGATTT AAACAGTTTA GTCAATAAAT ATAATCAAAT 2820 AGCCGGAAGCG AGACAGTGGC TCAGTGAATT GACCACTGA AATACCAGTA ATACCATGA 2760 AGCGGAAGCG AGACAGTGGC TCAGTGAATT GACCACTGA AATACCATTA 2940 AGCCGGAAGCG TAGGCGCCG GAAAAGACTC AAAACAGTTAC CAAAAAATGG TTCAGGATAT 2940 AGCCTGGCAG TCGGGTTTTA AAGCCGCCAA AGACACCGCA ATGGATAACG CCAAATACCA TACAGACACT AAAACCACAT TACAGACACT TACAGACACT AAAACCACAT TACAGACACT TACAGACACT TACAGACACT TACAGACACC CCAAATACC CCAAATACC AAAACCACAT TACAGACACC TACAGACACC CCAAATACC CCAAATACC AAAACCACAT TACAGACACC TACAGACACC CCAAATACC CCAAATACC AAAACCACAT TACAGACACC TACAGACACC CCAAATACC CCAAATACC CCAAATACC AAAACCACAT TACAGACACC TACAGACACC CCAAATACC CCAAATACC AAAACCACAT TACAGACACC TACAGACACC CCAAATACC CCAAATACC CCAAATACC CCAAATACC CCAAATACC CCAAATACC CCAAATACC AAAACCACAT TACAGACACC TACAGACACC CCAAATACC CCAAATACC CCAAATACC AAAACCACAT TACAGACCGC AAACCCCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3120 ATTAATAATA GGCAGCCAC CCAATTCATT GTACGACACC CTGGTAAAAC TGCTGAGCAC 3120 ATTAATAATA GGCAGCCAC CCAATTCATT GTACGACACC CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATA ACCGCCAAA AACCGCCAA AACCCCCCAC AAACCGCCAA AAGCTTCT CCACAGCACC GAAAAATAC AGAAGAGGAT 3240 ATTAATAATA ACCGCCAAACC CGCAATCTT TCCGCAGTCA GAGAAAGTGC 3300 CGAGATCAAA ACGCAGGCA CCGATCTT TCCGCAACTC GAGAAAATTC CCGCCCGTC ATAACCGCCAAACC CCGCCCGTC ATAACCGCCAACC CCGCCCGCC GAACCCCCGCC GAACCCCCGCC AAACCCCCAACCC GAGAAACTCT TCCGCCAGTCA GAGAAAGTCC CCCCCGTC ATAACCGCCAACC CCGCCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCG	CGCTTCTCCT		TCGTTGCCGA	ACGGCCGCAG	ACTCCCTCGG	CGAGCGAGCA	2280
ATCGCAGGCG GCTACTAAAA TCCACCAGGC ACAGCAGACG CTGCAGTCAA CGCCACCGAT 2400 CTCTGAAGAG AATAATGACG AGCGCACGCT GGCGCGCCAG CAGTTGACCA GCAGCCTGAA 2460 TGCGCTGGCG AAGTCCGGCG TGTCATTATC CGCAGAACAA AATGAGAACC TGCGGAGCGC 2520 GTTTTCTGCG CCGACGTCGG CCTTATTTAG CGCTTCGCCT ATGGCGCAGC CGAGAACAA 2580 CTATCTGGT GCTGAGATTT GGGATATGGT TTCCCAAAAT ATATCGGCGA TAGGTGACAG 2640 CTATCTGGC GTTTATGAAA ACGTTGTCC AGTCTATACC GATTTTTATC AGGCCTTCAG 2700 TGATATTCTT TCCAAAATGG GAGGCTGGTT ATTACCAGGT AAGGACGGTA ATACCGTTAA 2760 GCTAGATGATT ACCTCACTCA AAAATGATTT AAACAGTTTA GTCAATAAAT ATAATCAAAT 2820 AGACAGTAAA ACCGTTTATA TCCAGCGCA GTCAGGCAGC GGCGTTAAAG TAGCCACTGA 2880 TGGATCCGGT TATGTCGTCA CCGTTGATCT GACCCATTA CAAAAAATGG TTCAGGATAT 2940 AGCCTGGCAG TCGGGTTTTA AAGCGCAGA AGAACACCAT TACAGACGCT 3000 TGATGGTTTA GGCCGCCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3006 AGCCTGGCAG TCGGGTTTTA AAGCGCAGA AGAACACAC CTGGTAAAAG TGCTGACGG 3180 TACGATAAGT AGCAGCCTGG AAACCCCAA AAGCTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA AAACCCCCAA AAGCTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCA CAAATTAAAGG CCCCCCGTC ATAATCCCAG GAGAAAGTGC 3360 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCCCCCGTC ATAATTCC CCGCGCTGAT 3420	CGTCGAGACT		CGTCTACCAC	AGAACATCGC	GGTACAGATA	TCATTTCATT	2340
TGCGCTGGCG AAGTCCGGCG TGTCATTATC CGCAGAACAA AATGAGAACC TGCGGAGCGC 2520 GTTTTCTGCG CCGACGTCGG CCTTATTAG CGCTTCGCCT ATGGCGCAGC CGAGAACAAC 2580 CATTTCTGAT GCTGAGATTT GGGATATGGT TTCCCAAAAT ATATCGGCGA TAGGTGACAG 2640 CTATCTGGGC GTTTATGAAA ACGTTGTCGC AGTCTATACC GATTTTTATC AGGCCTTCAG 2700 TGATATTCTT TCCAAAATGG GAGGCTGGTT ATTACCAGGT AAGGACGGTA ATACCGTTAA 2760 GCTAGATGTT ACCTCACTCA AAAATGATTT AAACAGTTTA GTCAATAAAT ATAATCAAAT 2820 AAACAGTAAT ACCGTTTAT TTCCAGCGCA GTCAGGCAGC GGCGTTAAAG TAGCCACTGA 2880 AGCGGAAGCG AGACAGTGGC TCAGTGAATT GAATTTACCG AATAGCTGCC TGAAATCTTA 2940 TGATGGTTTA GGCGCCGCG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3000 TGATGGTTTA GGCGCCCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3060 AGCCTGGCAG TCGGGTTTTA AAGCGCCAGA AGAAAATATG AAAACCACAT TACAGACGCT 3120 GACGCAAAAA TATAGCAATG CCAATTCATT GTACGACAAC CTGGTAAAAG TGCTGAGCAG 3180 TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CAATTCTTGC GGCGAATCTT TCCGCAGTCA GAGAAAGTGC 3360 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTC CCGCGCTGAT 3420	ATCGCAGGCG	GCTACTAAAA	TCCACCAGGC	ACAGCAGACG	CTGCAGTCAA	CGCCACCGAT	2400
GTTTTCTGCG CCGACGTCGG CCTTATTTAG CGCTTCGCCT ATGGCGCAGC CGAGAACAAC 2580 CATTTCTGAT GCTGAGATTT GGGATATGGT TTCCCAAAAT ATATCGGCGA TAGGTGACAG 2640 CTATCTGGGC GTTTATGAAA ACGTTGTCGC AGTCTATACC GATTTTATC AGGCCTTCAG 2700 TGATATTCTT TCCAAAATGG GAGGCTGGTT ATTACCAGGT AAGGACGGTA ATACCGTTAA 2760 GCTAGATGTT ACCTCACTCA AAAATGATTT AAACAGTTTA GTCAATAAAT ATAATCAAAT 2820 AACAGTAAT ACCGTTTAT TTCCAGCGCA GTCAGGCAGC GGCGTTAAAG TAGCCACTGA 2880 AGCGGAAGCG AGACAGTGGC TCAGTGAATT GAATTTACCG AATAGCTGCC TGAAATCTTA 2940 TGATGGTTTA GGCGCCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3000 TGATGGTTTA GGCGCCCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3060 AGCCTGGCAG TCGGGTTTTA AAGCGCAGGA AGAAAATATG AAAACCACAT TACAGACGCT 3120 GACGCAAAAA TATAGCAATG CCAATTCATT GTACGACAAC CTGGTAAAAG TGCTGAGCAG 3180 TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CAAATTTAAAGG CCCGCAACTG GAAGATTTC CCGCGCTGAT 3420 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTC CCGCGCTGAT 3420	CTCTGAAGAG	AATAATGACG	AGCGCACGCT	GGCGCGCCAG	CAGTTGACCA	GCAGCCTGAA	2460
CATTTCTGAT GCTGAGATTT GGGATATGGT TTCCCAAAAT ATATCGGCGA TAGGTGACAG 2640 CTATCTGGGC GTTTATGAAA ACGTTGTCGC AGTCTATACC GATTTTTATC AGGCCTTCAG 2700 TGATATTCTT TCCAAAATGG GAGGCTGGTT ATTACCAGGT AAGGACGGTA ATACCGTTAA 2760 GCTAGATGTT ACCTCACTCA AAAATGATTT AAACAGTTTA GTCAATAAAT ATAATCAAAT 2820 AACAGTAAT ACCGTTTAT TTCCAGCGCA GTCAGGCAGC GGCGTTAAAG TAGCCACTGA 2880 AGCGGAAGCG AGACAGTGGC TCAGTGAATT GAATTTACCG AATAGCTGCC TGAAATCTTA 2940 TGATGGTTTA GGCGCCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3000 TGATGGTTTA GGCGCCCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3060 AGCCTGGCAG TCGGGTTTTA AAGCGCAGGA AGAAAATATG AAAACCACAT TACAGACGCT 3120 GACGCAAAAA TATAGCAATG CCAATTCATT GTACGACAAC CTGGTAAAAG TGCTGAGCAG 3180 TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CAAATTAAAGG CCCGCAACTG GAAGATTTC CCGCGCTGAT 3420	TGCGCTGGCG	AAGTCCGGCG	TGTCATTATC	CGCAGAACAA	AATGAGAACC	TGCGGAGCGC	2520
CTATCTGGGC GTTTATGAAA ACGTTGTCGC AGTCTATACC GATTTTATC AGGCCTTCAG 2700 TGATATTCTT TCCAAAATGG GAGGCTGGTT ATTACCAGGT AAGGACGGTA ATACCGTTAA 2760 GCTAGATGTT ACCTCACTCA AAAATGATTT AAACAGTTTA GTCAATAAAT ATAATCAAAT 2820 AAACAGTAAT ACCGTTTAT TTCCAGCGCA GTCAGGCAGC GGCGTTAAAG TAGCCACTGA 2880 AGCGGAAGCG AGACAGTGGC TCAGTGAATT GAATTTACCG AATAGCTGCC TGAAATCTTA 2940 TGGATCCGGT TATGTCGTCA CCGTTGATCT GACGCCATTA CAAAAAATGG TTCAGGATAT 3000 TGATGGTTTA GGCGCCCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3060 AGCCTGGCAG TCGGGTTTTA AAGCGCAGGA AGAAAATATG AAAACCACAT TACAGACGCT 3120 GACGCAAAAA TATAGCAATG CCAATTCATT GTACGACAAC CTGGTAAAAG TGCTGAGCAG 3180 TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CAAATTTAAAGG CCCGCAACTG GAAGATTTC CCGCGCTGAT 3420	GTTTTCTGCG	CCGACGTCGG	CCTTATTTAG	CGCTTCGCCT	ATGGCGCAGC	CGAGAACAAC	2580
TGATATTCTT TCCAAAATGG GAGGCTGGTT ATTACCAGGT AAGGACGGTA ATACCGTTAA 2760 GCTAGATGTT ACCTCACTCA AAAATGATTT AAACAGTTTA GTCAATAAAT ATAATCAAAT 2820 AAACAGTAAT ACCGTTTAT TTCCAGCGCA GTCAGGCAGC GGCGTTAAAG TAGCCACTGA 2880 AGCGGAAGCG AGACAGTGGC TCAGTGAATT GAATTTACCG AATAGCTGCC TGAAATCTTA 2940 TGATGGTTTA GGCGCCGCG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3000 TGATGGTTTA GGCGCCCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3060 AGCCTGGCAG TCGGGTTTTA AAGCGCAGGA AGAAAATATG AAAACCACAT TACAGACGCT 3120 GACGCAAAAA TATAGCAATG CCAATTCATT GTACGACAAC CTGGTAAAAG TGCTGAGCAG 3180 TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CGAATCTTGC GGCGAATCTT TCCGCAGTCA GAGAAAGTGC 3360 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTC CCGCGCTGAT 3420	CATTTCTGAT	GCTGAGATTT	GGGATATGGT	TTCCCAAAAT	ATATCGGCGA	TAGGTGACAG	2640
GCTAGATGTT ACCTCACTCA AAAATGATTT AAACAGTTTA GTCAATAAT ATAATCAAAT AAACAGTAAT ACCGTTTAT TTCCAGCGCA GTCAGGCAGC GGCGTTAAAG TAGCCACTGA 2880 AGCGGAAGCG AGACAGTGGC TCAGTGAATT GAATTTACCG AATAGCTGCC TGAAATCTTA 2940 TGGATCCGGT TATGTCGTCA CCGTTGATCT GACGCCATTA CAAAAAATGG TTCAGGATAT 3000 TGATGGTTTA GGCGCCCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3060 AGCCTGGCAG TCGGGTTTTA AAGCGCAGGA AGAAAATATG AAAACCACAT TACAGACGCT 3120 GACGCAAAAA TATAGCAATG CCAATTCATT GTACGACAAC CTGGTAAAAG TGCTGAGCAG 3180 TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CGAATCTTGC GGCGAATCTT TCCGCAGTCA GAGAAAGTGC 3360 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTC CCGCGCTGAT 3420	CTATCTGGGC	GTTTATGAAA	ACGTTGTCGC	AGTCTATACC	GATTTTTATC	AGGCCTTCAG	2700
AAACAGTAAT ACCGTTTTAT TTCCAGCGCA GTCAGGCAGC GGCGTTAAAG TAGCCACTGA 2880 AGCGGAAGCG AGACAGTGGC TCAGTGAATT GAATTTACCG AATAGCTGCC TGAAATCTTA 2940 TGGATCCGGT TATGTCGTCA CCGTTGATCT GACGCCATTA CAAAAAATGG TTCAGGATAT 3000 TGATGGTTTA GGCGCCCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3060 AGCCTGGCAG TCGGGTTTTA AAGCGCAGGA AGAAAATATG AAAACCACAT TACAGACGCT 3120 GACGCAAAAA TATAGCAATG CCAATTCATT GTACGACAAC CTGGTAAAAG TGCTGAGCAG 3180 TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CGAATCTTGC GGCGAATCTT TCCGCAGTCA GAGAAAGTGC 3360 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTC CCGCGCTGAT 3420	TGATATTCTT	TCCAAAATGG	GAGGCTGGTT	ATTACCAGGT	AAGGACGGTA	ATACCGTTAA	2760
AGCGGAAGCG AGACAGTGGC TCAGTGAATT GAATTTACCG AATAGCTGCC TGAAATCTTA 2940 TGATCCGGT TATGTCGTCA CCGTTGATCT GACGCCATTA CAAAAAATGG TTCAGGATAT 3000 TGATGGTTTA GGCGCCCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3060 AGCCTGGCAG TCGGGTTTTA AAGCGCAGGA AGAAAATATG AAAACCACAT TACAGACGCT 3120 GACGCAAAAA TATAGCAATG CCAATTCATT GTACGACAAC CTGGTAAAAG TGCTGAGCAG 3180 TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CGAATCTTGC GGCGAATCTT TCCGCAGTCA GAGAAAGTGC 3360 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTC CCGCGCTGAT 3420	GCTAGATGTT	ACCTCACTCA	AAAATGATTT	AAACAGTTTA	GTCAATAAAT	ATAATCAAAT	
TGGATCCGGT TATGTCGTCA CCGTTGATCT GACGCCATTA CAAAAAATGG TTCAGGATAT 3000 TGATGGTTTA GGCGCCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3060 AGCCTGGCAG TCGGGTTTTA AAGCGCAGGA AGAAAATATG AAAACCACAT TACAGACGCT 3120 GACGCAAAAA TATAGCAATG CCAATTCATT GTACGACAAC CTGGTAAAAG TGCTGAGCAG 3180 TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CGAATCTTGC GGCGAATCTT TCCGCAGTCA GAGAAAGTGC 3360 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTC CCGCGCTGAT 3420	AAACAGTAAT	ACCGTTTTAT	TTCCAGCGCA	GTCAGGCAGC	GGCGTTAAAG	TAGCCACTGA	2880
TGATGGTTTA GGCGCGCGG GAAAAGACTC AAAACTCGAA ATGGATAACG CCAAATATCA 3060 AGCCTGGCAG TCGGGTTTA AAGCGCAGGA AGAAAATATG AAAACCACAT TACAGACGCT 3120 GACGCAAAAA TATAGCAATG CCAATTCATT GTACGACAAC CTGGTAAAAG TGCTGAGCAG 3180 TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CGAATCTTGC GGCGAATCTT TCCGCAGTCA GAGAAAGTGC 3360 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTC CCGCGCTGAT 3420	AGCGGAAGCG	AGACAGTGGC	TCAGTGAATT	GAATTTACCG	AATAGCTGCC	TGAAATCTTA	
AGCCTGGCAG TCGGGTTTTA AAGCGCAGGA AGAAAATATG AAAACCACAT TACAGACGCT 3120 GACGCAAAAA TATAGCAATG CCAATTCATT GTACGACAAC CTGGTAAAAG TGCTGAGCAG TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CGAATCTTGC GGCGAATCTT TCCGCAGTCA GAGAAAGTGC 3360 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTTC CCGCGCTGAT 3420	TGGATCCGGT	TATGTCGTCA	CCGTTGATCT	GACGCCATTA	CAAAAAATGG	TTCAGGATAT	
GACGCAAAAA TATAGCAATG CCAATTCATT GTACGACAAC CTGGTAAAAG TGCTGAGCAG 3180 TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT 3240 ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CGAATCTTGC GGCGAATCTT TCCGCAGTCA GAGAAAGTGC 3360 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTTC CCGCGCTGAT 3420	TGATGGTTTA	GGCGCGCCGG	GAAAAGACTC	AAAACTCGAA	ATGGATAACG	CCAAATATCA	3060
TACGATAAGT AGCAGCCTGG AAACCGCCAA AAGCTTCCTG CAAGGATAAC AGAAGAGGAT ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCCGTC ATAATGCCAG GTATGCAGAC CGAGATCAAA ACGCAGGCCA CGAATCTTGC GGCGAATCTT TCCGCAGTCA GAGAAAGTGC CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTTC CCGCGCTGAT 3240 3360 3360	AGCCTGGCAG	TCGGGTTTTA	AAGCGCAGGA	AGAAAATATG	AAAACCACAT	TACAGACGCT	
ATTAATAATG GTTACAAGTG TAAGGACTCA GCCCCCGTC ATAATGCCAG GTATGCAGAC 3300 CGAGATCAAA ACGCAGGCCA CGAATCTTGC GGCGAATCTT TCCGCAGTCA GAGAAAGTGC 3360 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTTC CCGCGCTGAT 3420	GACGCAAAAA	TATAGCAATG	CCAATTCATT	GTACGACAAC	CTGGTAAAAG	TGCTGAGCAG	3180
CGAGATCAAA ACGCAGGCCA CGAATCTTGC GGCGAATCTT TCCGCAGTCA GAGAAAGTGC 3360 CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTTC CCGCGCTGAT 3420	TACGATAAGT	AGCAGCCTGG	AAACCGCCAA	AAGCTTCCTG	CAAGGATAAC	AGAAGAGGAT	
CACAGCGACG CTGTCAGGGG AAATTAAAGG CCCGCAACTG GAAGATTTTC CCGCGCTGAT 3420	ATTAATAATG	GTTACAAGTG	TAAGGACTCA	GCCCCCGTC	ATAATGCCAG	GTATGCAGAC	
0110110001100 01010110000 1221121101 0000121010 0121111111 011111111	CGAGATCAAA	ACGCAGGCCA	CGAATCTTGC	GGCGAATCTT	TCCGCAGTCA	GAGAAAGTGC	
2444	CACAGCGACG			CCCGCAACTG	GAAGATTTTC	CCGCGCTGAT	
CAAACAGGCG AGTCTGGATG C 3441	CAAACAGGCG	AGTCTGGATG	C				3441

# (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 3851 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: double(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GAGCTCAGCA	ACGTGTCGAA	AGCCTGTAAA	ATCATGGGCG	TCTCGCGCGA	TACGTTTTAC	60
CGTTATCGTG	<b>AACTGGCCGA</b>	TGAAGGCGGC	GTTGATGCGC	TGATAAATCG	TAGTCGCCGC	120
GTACCTAACC	TTAAGAACCG	TACCGATGAG	GCAACTGAGC	AAGCTGTTGT	TGATTATGCC	180
GTTGCGTTCC	CGGCCCATGG	TCAGCACCGA	ACTGCGCAAA	CAGGACGTTT	TTATCTCCGG	240
TAGTGATGTC	CATTCCGTCT	GGCTGCGCAC	<b>AACCTTGAGA</b>	ACTTCAAAAA	ACGCCTGAAA	300
GCGCTGGAAG	AAAAAGTGGC	CCGCGATGGC	ATTGAACTGA	CTGCCAGATC	GCCGCGCTGG	360
AGCGTAAAGC	CAGTGATGAT	GAAGCCTGTG	GTGAGATTGA	AACCGTTCAT	CCGGGATATC	420
TGGGGTCACA	GGACACGTTC	TACGTGGGCA	ACCTGAAAGG	CGTTGGGCGA	ATCTATCAGC	480
AGACGTTCGT	TGATACATAC	TCGAAGGTGG	CTCACTGCAA	GCGCTATATC	ACCAAAACGC	540
CGATTACAGC	GGCTGATTTG	CTGAATGATC	GTGTACTGCC	GTTTATGAGT	CTCAGGGCCT	600
GCCGATGCTA	AGGATACTGA	CAGACAGGGG	TACAGAATAT	TGCGGCAAAG	TGGAACATCA	660
TGATTATCAG	CTTTATCTGG	AGATAAATGA	CATCGAACAC	ACGAAAACGA	AGGCGATGTC	720

CCCGCAGACC	AATGGCATCT	GCGAGCGGTT	CCATAAAACG	ATACTGAACG	AATTTTATCA	780
GGTGACGTTC	CGCAAAAAGT	TATATGGCGA	TTTTGATACA	TTACAATCGG	ATCTTGATGA	840
			CCATCAGGGA			900
GCCGATGGAA	ACGTTACTTG	ATGGAAAACG	CATCTGGTCT	GAGAAGAATT	TAAGCCAGAT	960
			GGTAACTGTC		TGAGCTAATA	1020
CAACTAATTG			GCTAAATATA			1080
	CTGCGTTAAT		TGTGCTGTAA	ATTAGGCAGT	GGAATATGTT	1140
	AATACACAAC		TATGCAGGCT	ATTGCTGGTG	CAGCGGCACC	1200
			AAAATTCAGG	TTTTTTTCCC	GCAGGAAAAT	1260
			CTTTGCCATC	CGGCCCGGGG	AATGTTGCCT	1320
	GCAGCCAGTT		AAAGCGCTTA			1380
AATATTCAGT	• • • • • • • • • • • • • • • • • • • •		CAGTTTTGTA			1440
	CAATCACTCT		GGGAACTATA			1500
AGTGAAGCAC			GAACCGGGAG			1560
			GCCACAACGG			1620
			ATGCACCAGC			1680
			TGAATAACGG			1740
GAGCGTCAGG			GTTTACCACC			1800
			AGTTGCCGGA			1860
			CATCGCTGCC			1920
			CGACGTTGCC			1980
			CGGAGATGCC			2040
	TAACCAGCTG		CGCGCTGCCG			2100
			GGAGATGTCG			2160
			TGCTTAACGT			2220
	CCTGAATAAC				AATAATCTGA	2280
		CCGCATATTA				2340
			GGGAGCTGGA			2400
	GCCATCGCTG		TACAGAAGCT		AATAATTGGC	
			TAGGGGATCT			2460
			TACGGGAGCT			2520
		CGTCAGGACT			ATAATCGGCT	2580
			ACAGGAGCTG			2640
	CCGCAAAGCC		GTCTTCAGCG			2700
			GCTCTGCGGA			2760
	CTTCGATATG		CCGTCCCCGG			2820
			GGAGGGGGAA			2880
			CGCCTTCAGC			2940
			CTTTAAGGCA			3000
			AAAAACCTTT			3060
			CCTGCACCAG			3120
			CAATCTCCAG			3180
			GATTGCCCGG			3240
			GTTCCAGAAT	AAGCTGAAGG	AATCACTTGA	3300
GCTGACCAGC	GTGACGTCAG	AAATGCGTTT	CTTTGACGTT	TCCGGCGTGA	CGGTTTCAGA	3360
CCTTCAGGCT	GCGGACGTTC	AGGTGAAAAC	CGCTGAAAAC	AGCGGGTTCA	GTAAATGGAT	3420
ACTGCAGTGG	GGGCCGTTAC	ACAGCGTGCT	GGAACGCAAA	GTGCCGGAAC	GCTTTAACGC	3480
GCTTCGTGAA	<b>AAGCAAATAT</b>	CGGATTATGA	AGACACGTAC	CGGAAGCTGT	ATGACGAAGT	3540
GCTGAAATCG	TCCGGGCTGG	TCGACGATAC	CGATGCAGAA	CGTACTATCG	GAGTAAGTGC	3600
GATGGATAGT	GCGAAAAAAG	<b>AATTTCTGGA</b>	TGGCCTGCGC	GCTCTTGTGG	ATGAGGTGCT	3660
GGGTAGCTAT	CTGACAGCCC	GGTGGCGTCT	TAACTGAGCA	CGATATTCTC	CGCACCAGGC	3720
GAATGTGGTG	CGGTGAACAA	AGATATTCCT	TGGACAAACA	<b>ACATGAGACA</b>	GCACTGATGA	3780
			GTCAGGGCGT			3840
CGATACGCGC						3851

# (2) INFORMATION FOR SEQ ID NO:17:

# (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: double
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:	
CCAGGTGGAT ACGGA	15
(2) INFORMATION FOR SEQ ID NO:18:	
(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: double  (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:	
TAGCGTCCTC CCCATGTGCG	20
(2) INFORMATION FOR SEQ ID NO:19:	
<ul><li>(i) SEQUENCE CHARACTERISTICS:</li><li>(A) LENGTH: 18 base pairs</li><li>(B) TYPE: nucleic acid</li><li>(C) STRANDEDNESS: double</li><li>(D) TOPOLOGY: linear</li></ul>	
(ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:	
CCGGCGCTAC TGGCGGCG	18
(2) INFORMATION FOR SEQ ID NO:20:	
<ul><li>(i) SEQUENCE CHARACTERISTICS:</li><li>(A) LENGTH: 18 base pairs</li><li>(B) TYPE: nucleic acid</li><li>(C) STRANDEDNESS: double</li><li>(D) TOPOLOGY: linear</li></ul>	
(ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:	
AGCGTTTCAA CAGCCCCG	18
(2) INFORMATION FOR SEQ ID NO:21:	
<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 17 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: double</li> <li>(D) TOPOLOGY: linear</li> </ul>	
(ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:	
CGGGGCTGTT GAAACGC	17

(2) INFORMATION FOR SEQ ID NO:22:	
<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 16 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: double</li> <li>(D) TOPOLOGY: linear</li> </ul>	
(ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:	
AACCTGGCCT TTTCAG	16
(2) INFORMATION FOR SEQ ID NO:23:	
<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 19 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: double</li> <li>(D) TOPOLOGY: linear</li> </ul>	
(ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:	
GGCAGGGAGC CTTGCTTGG	19
(2) INFORMATION FOR SEQ ID NO:24:	
<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 19 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: double</li> <li>(D) TOPOLOGY: linear</li> </ul>	
(ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:	
GTGCCTGGCC AGTTCTCCA	19
(2) INFORMATION FOR SEQ ID NO:25:	
<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 26 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: double</li> <li>(D) TOPOLOGY: linear</li> </ul>	
(ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:	
CGCGGATCCA TTATGGGATG TATCGG	26
(2) INFORMATION FOR SEQ ID NO:26:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid	

(C) STRANDEDNESS: double (D) TOPOLOGY: linear (ii) MOLECULE TYPE: Genomic DNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26: CCGGCAGCAA AATGTTGCAG (2) INFORMATION FOR SEQ ID NO:27: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 12 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear (ii) MOLECULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27: Val Thr Ser Val Arg Thr Gln Pro Pro Val Ile Met (2) INFORMATION FOR SEQ ID NO:28: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 15 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear (ii) MOLECULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28: Met Leu Ile Ser Asn Val Gly Ile Asn Pro Ala Ala Tyr Leu Asn 5 (2) INFORMATION FOR SEQ ID NO:29: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 13 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear (ii) MOLECULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29: Met Asn Asn Leu Thr Leu Ser Xaa Phe Xaa Lys Val Gly (2) INFORMATION FOR SEQ ID NO:30: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 15 amino acids

1

20

(B) TYPE: amino acid (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Met Leu Ile Ser Asn Val Gly Ile Asn Pro Ala Ala Tyr Leu Asn 1 5 10 15

- (2) INFORMATION FOR SEQ ID NO:31:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 12 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Thr Lys Ile Thr Leu Ser Pro Gln Asn Phe Phe Ile
1 5 10

- (2) INFORMATION FOR SEQ ID NO:32:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 80 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

 Met
 Ala
 Thr
 Pro
 Trp
 Ser
 Gly
 Tyr
 Leu
 Asp
 Asp
 Val
 Ser
 Ala
 Lys
 Leu
 Asp
 Asp
 Asp
 Leu
 Gln
 Thr
 Gln
 Val
 Thr
 Glu
 Ala
 Leu
 Asp
 Asp
 Asp
 Pro
 Ala
 Leu
 Leu
 Ala
 Ala
 Ala
 Tyr
 Gln
 Asp
 Asp
 Asp
 Ala
 Ala
 Gln
 Ser
 Asp
 Asp
 Ala
 Ala
 Ala
 Gln
 Ser
 Asp
 Asp
 Ala
 Ala
 Ile
 I

- (2) INFORMATION FOR SEQ ID NO:33:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 83 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Ser Asn Thr Val Lys Val Ile Lys Asp Val Asp Ala Ala Ile Ile Gln 65 70 75 80 Asn Phe Arg

- (2) INFORMATION FOR SEQ ID NO:34
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 87 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

 Met
 Ser
 Asn
 Phe
 Ser
 Gly
 Phe
 Thr
 Lys
 Gly
 Asn
 Asp
 Ile
 Ala
 Asp
 Leu

 Asp
 Ala
 Val
 Ala
 Gln
 Thr
 Leu
 Lys
 Lys
 Pro
 Ala
 Asp
 Asp
 Ala
 Lys
 Asp
 A

- (2) INFORMATION FOR SEQ ID NO:35:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 101 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Met Ser Ile Ala Thr Ile Val Pro Glu Asn Ala Val Ile Gly Gln Ala 10 Val Asn Ile Arg Ser Met Glu Thr Asp Ile Val Ser Leu Asp Asp Arg 25 Leu Leu Gln Ala Phe Ser Gly Ser Ala Ile Ala Thr Ala Val Asp Lys 40 35 Gln Thr Ile Thr Asn Arg Ile Glu Asp Pro Asn Leu Val Thr Asp Pro 55 Lys Glu Leu Ala Ile Ser Gln Glu Met Ile Ser Asp Tyr Asn Leu Tyr 75 70 Val Ser Met Val Ser Thr Leu Thr Arg Lys Gly Val Gly Ala Val Glu 85 Thr Leu Leu Arg Ser 100

- (2) INFORMATION FOR SEQ ID NO:36:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 97 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

 Met
 Asn
 Tyr
 Ile
 Tyr
 Pro
 Val
 Asn
 Gln
 Val
 Asp
 Ile
 Ile
 Lys
 Ala
 Ser

 Asp
 Phe
 Gln
 Ser
 Glu
 Ile
 Ser
 Ser
 Leu
 Glu
 Asp
 Val
 Ser
 Ala

 Lys
 Tyr
 Ser
 Asp
 Ile
 Lys
 Met
 Asp
 Thr
 Asp
 Ile
 Gln
 Val
 Ser
 Gln
 Ile

 Met
 Glu
 Met
 Val
 Ser
 Asp
 Pro
 Glu
 Ser
 Leu
 Asp
 Pro
 Glu
 Ser
 Leu
 Asp
 Pro
 Glu
 Ser
 Leu
 Ala
 Ala
 Ser
 Leu
 Ala
 Ala
 Ser
 Leu
 Ala
 Ala
 Val
 Ser
 Leu
 Ala
 A

### (2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 241 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Met Ile Arg Tyr Lys Gly Phe Ile Leu Phe Leu Leu Met Leu Ile 10 Gly Cys Glu Gln Arg Glu Glu Leu Ile Ser Asn Leu Ser Gln Arg Gln Ala Asn Glu Ile Ile Ser Val Leu Glu Arg His Asn Ile Thr Ala Arg 40 Lys Val Asp Gly Gly Lys Gln Gly Ile Ser Val Gln Val Glu Lys Gly Thr Phe Ala Ser Ala Val Asp Leu Met Arg Met Tyr Asp Leu Pro Asn 70 75 Pro Glu Arg Val Asp Ile Ser Gln Met Phe Pro Thr Asp Ser Leu Val 85 90 Ser Ser Pro Arg Ala Glu Lys Ala Arg Leu Tyr Ser Ala Ile Glu Gln 100 105 Arg Leu Glu Gln Ser Leu Val Ser Ile Gly Gly Val Ile Ser Ala Lys 120 125 Ile His Val Ser Tyr Asp Leu Glu Glu Lys Asn Ile Ser Ser Lys Pro 135 140 Met His Ile Ser Val Ile Ala Ile Tyr Asp Ser Pro Lys Glu Ser Glu 150 155 Leu Leu Val Ser Asn Ile Lys Arg Phe Leu Lys Asn Thr Phe Ser Asp 170 165 175 Val Lys Tyr Glu Asn Ile Ser Val Ile Leu Thr Pro Lys Glu Glu Tyr 185 180 Val Tyr Thr Asn Val Gln Pro Val Lys Glu Val Lys Ser Glu Phe Leu 200 205 Thr Asn Glu Val Ile Tyr Leu Phe Leu Gly Met Ala Val Leu Val Val 215 220 Ile Leu Leu Val Trp Ala Phe Lys Thr Gly Trp Phe Lys Arg Asn Lys 235 225 230 Ile

## (2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 244 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Met Lys Val Lys Thr Ser Leu Ser Thr Leu Ile Leu Ile Leu Phe Leu 10 Thr Gly Cys Lys Val Asp Leu Tyr Thr Gly Ile Ser Gln Lys Glu Gly 30 20 25 Asn Glu Met Leu Ala Leu Leu Arg Gln Glu Gly Leu Ser Ala Asp Lys 40 Glu Pro Asp Lys Asp Gly Lys Ile Lys Leu Leu Val Glu Glu Ser Asp 55 Val Ala Gln Ala Ile Asp Ile Leu Lys Arg Lys Gly Tyr Pro His Glu 75 70 Ser Phe Ser Thr Leu Gln Asp Val Phe Pro Lys Asp Gly Leu Ile Ser 85 90 Ser Pro Ile Glu Glu Leu Ala Arg Leu Asn Tyr Ala Lys Ala Gln Glu 100 105 110 Ile Ser Arg Thr Leu Ser Glu Ile Asp Gly Val Leu Val Ala Arg Val 120 125 115 His Val Val Leu Pro Glu Glu Gln Asn Asn Lys Gly Lys Lys Gly Val 135 140 Ala Ala Ser Ala Ser Val Phe Ile Lys His Ala Ala Asp Ile Gln Phe 155 150 Asp Thr Tyr Ile Pro Gln Ile Lys Gln Leu Val Asn Asn Ser Ile Glu 165 170 Gly Leu Ala Tyr Asp Arg Ile Ser Val Ile Leu Val Pro Ser Val Asp 180 185 190 **Val Arg Gln Ser Ser His Leu Pro Arg Asn Thr Ser Ile Leu Ser Ile** 200 205 195 Gln Val Ser Glu Glu Ser Lys Gly Arg Leu Ile Gly Leu Leu Ser Leu 215 220 Leu Ile Leu Leu Pro Val Thr Asn Leu Ala Gln Tyr Phe Trp Leu 235 230 Gln Arg Lys Lys

## (2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 252 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

 Met Ile Arg Arg Tyr Leu Tyr Thr Phe Leu Leu Val Met Thr Leu Ala

 1
 5
 10
 15

 Gly Cys Lys Asp Lys Asp Leu Leu Lys Gly Leu Asp Gln Glu Gln Ala
 20
 25
 30

 Asn Glu Val Ile Ala Val Leu Gln Met His Asn Ile Glu Ala Asn Lys
 35
 40
 45

Ile Asp Ser Gly Lys Leu Gly Tyr Ser Ile Thr Val Ala Glu Pro Asp Phe Thr Ala Ala Val Tyr Trp Ile Lys Thr Tyr Gln Leu Pro Pro Arg 70 75 Pro Arg Val Glu Ile Ala Gln Met Phe Pro Ala Asp Ser Leu Val Ser 85 90 Ser Pro Arg Ala Glu Lys Ala Arg Leu Tyr Ser Ala Ile Glu Gln Arg 105 110 Leu Glu Gln Ser Leu Gln Thr Met Glu Gly Val Leu Ser Ala Arg Val 120 125 115 His Ile Ser Tyr Asp Ile Asp Ala Gly Glu Asn Gly Arg Pro Pro Lys 135 Pro Val His Leu Ser Ala Leu Ala Val Tyr Glu Arg Gly Ser Pro Leu 150 155 Ala His Gln Ile Ser Asp Ile Lys Arg Phe Leu Lys Asn Ser Phe Ala 170 165 Asp Val Asp Tyr Asp Asn Ile Ser Val Val Leu Ser Glu Arg Ser Asp 185 190 Ala Gln Leu Gln Ala Pro Gly Thr Pro Val Lys Arg Asn Ser Phe Ala 200 205 195 Thr Ser Trp Ile Val Leu Ile Ile Leu Leu Ser Val Met Ser Ala Gly 215 220 210 Phe Gly Val Trp Tyr Tyr Lys Asn His Tyr Ala Arg Asn Lys Lys Gly 230 235 Ile Thr Ala Asp Asp Lys Ala Lys Ser Ser Asn Glu 245

## (2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 289 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Arg Lys Ala Glu Glu Thr Asn Arg Ile Met Gly Cys Ile Gly Lys Val 10 Leu Gly Ala Leu Leu Thr Ile Val Ser Val Val Ala Ala Val Phe Thr 25 20 Gly Gly Ala Ser Leu Ala Leu Ala Ala Val Gly Leu Ala Val Met Val 40 Ala Asp Glu Ile Val Lys Ala Ala Thr Gly Val Ser Phe Ile Gln Gln 55 Ala Leu Asn Pro Ile Met Glu His Val Leu Lys Pro Leu Met Glu Leu 70 Ile Gly Lys Ala Ile Thr Lys Ala Leu Glu Gly Leu Gly Val Asp Lys 90 Lys Thr Ala Glu Met Ala Gly Ser Ile Val Gly Ala Ile Val Ala Ala 105 100 Ile Ala Met Val Ala Val Ile Val Val Val Ala Val Val Gly Lys Gly 120 Ala Ala Ala Lys Leu Gly Asn Ala Leu Ser Lys Met Met Gly Glu Thr 135 140 Ile Lys Lys Leu Val Pro Asn Val Leu Lys Gln Leu Ala Gln Asn Gly 150 155 Ser Lys Leu Phe Thr Gln Gly Met Gln Arg Ile Thr Ser Gly Leu Gly 170 175 Asn Val Gly Ser Lys Met Gly Leu Gln Thr Asn Ala Leu Ser Lys Glu 180 185

Leu Val Gly Asn Thr Leu Asn Lys Val Ala Leu Gly Met Glu Val Thr 200 195 Asn Thr Ala Ala Gln Ser Ala Gly Gly Val Ala Glu Gly Val Phe Ile 215 220 Lys Asn Ala Ser Glu Ala Leu Ala Asp Phe Met Leu Ala Arg Phe Ala 235 230 Met Asp Gln Ile Gln Gln Trp Leu Lys Gln Ser Val Glu Ile Phe Gly 245 250 Glu Asn Gln Lys Val Thr Ala Glu Leu Gln Lys Ala Met Ser Ser Ala 265 270 260 Val Gln Gln Asn Ala Asp Ala Ser Arg Phe Ile Leu Arg Gln Ser Arg 280 275 Ala

# (2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 283 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Arg Lys Ala Glu Glu Leu Asn Arg Val Met Gly Cys Val Gly Lys Ile 15 Leu Gly Ala Leu Leu Thr Ile Val Ser Val Val Ala Ala Ala Phe Ser 20 25 Gly Gly Ala Ser Leu Ala Leu Ala Ala Val Gly Leu Ala Leu Met Val 40 Thr Asp Ala Ile Val Gln Ala Ala Thr Gly Asn Ser Phe Met Glu Gln Ala Leu Asn Pro Ile Met Lys Ala Val Ile Glu Pro Leu Ile Lys Leu 70 75 Leu Ser Asp Ala Phe Thr Lys Met Leu Glu Gly Leu Gly Val Asp Ser 90 Lys Lys Ala Lys Met Ile Gly Ser Ile Leu Gly Ala Ile Ala Gly Ala 105 100 110 Leu Val Leu Val Ala Ala Val Val Leu Val Ala Thr Val Gly Lys Gln 125 115 120 Ala Ala Ala Lys Leu Ala Glu Asn Ile Gly Lys Ile Ile Gly Lys Thr 135 140 Leu Thr Asp Leu Ile Pro Lys Phe Leu Lys Asn Phe Ser Ser Gln Leu 150 155 Asp Asp Leu Ile Thr Asn Ala Val Ala Arg Leu Asn Lys Phe Leu Gly 170 175 165 Ala Ala Gly Asp Glu Val Ile Ser Lys Gln Ile Ile Ser Thr His Leu 185 180 Asn Gln Ala Val Leu Leu Gly Glu Ser Val Asn Ser Ala Thr Gln Ala 200 195 205 Gly Gly Ser Val Ala Ser Ala Val Phe Gln Asn Ser Ala Ser Thr Asn 210 215 220 Leu Ala Asp Leu Thr Leu Ser Lys Tyr Gln Val Glu Gln Leu Ser Lys 235 230 Tyr Ile Ser Glu Ala Ile Glu Lys Phe Gly Gln Leu Gln Glu Val Ile 245 250 Ala Asp Leu Leu Ala Ser Met Ser Asn Ser Gln Ala Asn Arg Thr Asp 265 260 Val Ala Lys Ala Ile Leu Gln Gln Thr Thr Ala 280 275

# (2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 409 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Met Leu Ile Ser Asn Val Gly Ile Asn Pro Ala Ala Tyr Leu Asn Asn His Ser Val Glu Asn Ser Ser Gln Thr Ala Ser Gln Ser Val Ser Ala 25 20 Lys Asp Ile Leu Asn Ser Ile Gly Ile Ser Ser Ser Lys Val Ser Asp 40 Leu Gly Leu Ser Pro Thr Leu Ser Ala Pro Ala Pro Gly Val Leu Thr 55 Gln Thr Pro Gly Thr Ile Thr Ser Phe Leu Lys Ala Ser Ile Gln Asn 70 Thr Asp Met Asn Gln Asp Leu Asn Ala Leu Ala Asn Asn Val Thr Thr 90 Lys Ala Asn Glu Val Val Gln Thr Gln Leu Arg Glu Gln Gln Ala Glu 110 105 100 Val Gly Lys Phe Phe Asp Ile Ser Gly Met Ser Ser Ser Ala Val Ala 120 125 115 Leu Leu Ala Ala Ala Asn Thr Leu Met Leu Thr Leu Asn Gln Ala Asp 135 140 Ser Lys Leu Ser Gly Lys Leu Ser Leu Val Ser Phe Asp Ala Ala Lys 150 155 Thr Thr Ala Ser Ser Met Met Arg Glu Gly Met Asn Ala Leu Ser Gly 170 165 Ser Ile Ser Gln Ser Ala Leu Gln Leu Gly Ile Thr Gly Val Gly Ala 180 185 Lys Leu Glu Tyr Lys Gly Leu Gln Asn Glu Arg Gly Ala Leu Lys His 200 Asn Ala Ala Lys Ile Asp Lys Leu Thr Thr Glu Ser His Ser Ile Lys 215 220 Asn Val Leu Asn Gly Gln Asn Ser Val Lys Leu Gly Ala Glu Gly Val 235 230 Asp Ser Leu Lys Ser Leu Asn Met Lys Lys Thr Gly Thr Asp Ala Thr 250 255 245 Lys Asn Leu Asn Asp Ala Thr Leu Lys Ser Asn Ala Gly Thr Ser Ala 265 270 260 Thr Glu Ser Leu Gly Ile Lys Asp Ser Asn Lys Gln Ile Ser Pro Glu 280 285 275 His Gln Ala Ile Leu Ser Lys Arg Leu Glu Ser Val Glu Ser Asp Ile 295 300 Arg Leu Glu Gln Asn Thr Met Asp Met Thr Arg Ile Asp Ala Arg Lys 310 315 Met Gln Met Thr Gly Asp Leu Ile Met Lys Asn Ser Val Thr Val Gly 325 330 Gly Ile Ala Gly Ala Ser Gly Gln Tyr Ala Ala Thr Gln Glu Arg Ser 340 345 Glu Gln Gln Ile Ser Gln Val Asn Asn Arg Val Ala Ser Thr Ala Ser 360 365 355 Asp Glu Ala Arg Glu Ser Ser Arg Lys Ser Thr Ser Leu Ile Gln Glu 375 380 Met Leu Lys Thr Met Glu Ser Ile Asn Gln Ser Lys Ala Ser Ala Leu 395 390

### Ala Ala Ile Ala Gly Asn Ile Arg Ala 405

# (2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 382 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Met Leu Gln Lys Gln Phe Cys Asn Lys Leu Leu Asp Thr Asn Lys 10 Glu Asn Val Met Glu Ile Gln Asn Thr Lys Pro Thr Gln Thr Leu Tyr 20 25 Thr Asp Ile Ser Thr Lys Gln Thr Gln Ser Ser Ser Glu Thr Gln Lys 40 Ser Gln Asn Tyr Gln Gln Ile Ala Ala His Ile Pro Leu Asn Val Gly 55 Lys Asn Pro Val Leu Thr Thr Leu Asn Asp Asp Gln Leu Leu Lys 70 Leu Ser Glu Gln Val Gln His Asp Ser Glu Ile Ile Ala Arg Leu Thr 90 85 Asp Lys Lys Met Lys Asp Leu Ser Glu Met Ser His Thr Leu Thr Pro 105 110 100 Glu Asn Thr Leu Asp Ile Ser Ser Leu Ser Ser Asn Ala Val Ser Leu 120 125 Ile Ile Ser Val Ala Val Leu Leu Ser Ala Leu Arg Thr Ala Glu Thr 140 135 Lys Leu Gly Ser Gln Leu Ser Leu Ile Ala Phe Asp Ala Thr Lys Ser 155 150 Ala Ala Glu Asn Ile Val Arg Gln Gly Leu Ala Ala Leu Ser Ser Ser 165 170 Ile Thr Gly Ala Val Thr Gln Val Gly Ile Thr Gly Ile Gly Ala Lys 185 180 Lys Thr His Ser Gly Ile Ser Asp Gln Lys Gly Ala Leu Arg Lys Asn 200 205 Leu Ala Thr Ala Gln Ser Leu Glu Lys Glu Leu Ala Gly Ser Lys Leu 210 215 220 Gly Leu Asn Lys Gln Ile Asp Thr Asn Ile Thr Ser Pro Gln Thr Asn 235 230 Ser Ser Thr Lys Phe Leu Gly Lys Asn Lys Leu Ala Pro Asp Asn Ile 245 250 255 Ser Leu Ser Thr Glu His Lys Thr Ser Leu Ser Ser Pro Asp Ile Ser 265 270 260 Leu Gln Asp Lys Ile Asp Thr Gln Arg Arg Thr Tyr Glu Leu Asn Thr 280 285 275 Leu Ser Ala Gln Gln Lys Gln Asn Ile Gly Arg Ala Thr Met Glu Thr 300 295 Ser Ala Val Ala Gly Asn Ile Ser Thr Ser Gly Gly Arg Tyr Ala Ser 315 310 Ala Leu Glu Glu Glu Gln Leu Ile Ser Gln Ala Ser Ser Lys Gln 325 330 335 Ala Glu Glu Ala Ser Gln Val Ser Lys Glu Ala Ser Gln Ala Thr Asn 345 340 Gln Leu Ile Gln Lys Leu Leu Asn Ile Ile Asp Ser Ile Asn Gln Ser 360 Lys Asn Ser Ala Ala Ser Gln Ile Ala Gly Asn Ile Arg Ala 375

# (2) INFORMATION FOR SEQ ID NO:44:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 343 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Met Leu Asn Ile Gln Asn Tyr Ser Ala Ser Pro His Pro Gly Ile Val 1 Ala Glu Arg Pro Gln Thr Pro Ser Ala Ser Glu His Val Glu Thr Ala 30 25 20 Val Val Pro Ser Thr Thr Glu His Arg Gly Thr Asp Ile Ile Ser Leu 40 35 Ser Gln Ala Ala Thr Lys Ile His Gln Ala Gln Gln Thr Leu Gln Ser 55 60 Thr Pro Pro Ile Ser Glu Glu Asn Asn Asp Glu Arg Thr Leu Ala Arg 70 75 Gln Gln Leu Thr Ser Ser Leu Asn Ala Leu Ala Lys Ser Gly Val Ser 85 90 Leu Ser Ala Glu Gln Asn Glu Asn Leu Arg Ser Ala Phe Ser Ala Pro 110 100 105 Thr Ser Ala Leu Phe Ser Ala Ser Pro Met Ala Gln Pro Arg Thr Thr 120 125 115 Ile Ser Asp Ala Glu Ile Trp Asp Met Val Ser Gln Asn Ile Ser Ala 135 140 Ile Gly Asp Ser Tyr Leu Gly Val Tyr Glu Asn Val Val Ala Val Tyr 150 155 Thr Asp Phe Tyr Gln Ala Phe Ser Asp Ile Leu Ser Lys Met Gly Gly 170 175 165 Trp Leu Leu Pro Gly Lys Asp Gly Asn Thr Val Lys Leu Asp Val Thr 190 180 185 Ser Leu Lys Asn Asp Leu Asn Ser Leu Val Asn Lys Tyr Asn Gln Ile 205 200 195 Asn Ser Asn Thr Val Leu Phe Pro Ala Gln Ser Gly Ser Gly Val Lys 215 220 Val Ala Thr Glu Ala Glu Ala Arg Gln Trp Leu Ser Glu Leu Asn Leu 235 230 Pro Asn Ser Cys Leu Lys Ser Tyr Gly Ser Gly Tyr Val Val Thr Val 255 250 245 Asp Leu Thr Pro Leu Gln Lys Met Val Gln Asp Ile Asp Gly Leu Gly 265 270 260 Ala Pro Gly Lys Asp Ser Lys Leu Glu Met Asp Asn Ala Lys Tyr Gln 280 285 275 Ala Trp Gln Ser Gly Phe Lys Ala Gln Glu Glu Asn Met Lys Thr Thr 290 295 Leu Gln Thr Leu Thr Gln Lys Tyr Ser Asn Ala Asn Ser Leu Tyr Asp 320 310 315 Asn Leu Val Lys Val Leu Ser Ser Thr Ile Ser Ser Ser Leu Glu Thr 330 325 Ala Lys Ser Phe Leu Gln Gly 340

### (2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 332 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: peptide

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Met Asn Ile Thr Thr Leu Thr Asn Ser Ile Ser Thr Ser Ser Phe Ser 10 Pro Asn Asn Thr Asn Gly Ser Ser Thr Glu Thr Val Asn Ser Asp Ile 20 25 Lys Thr Thr Ser Ser His Pro Val Ser Ser Leu Thr Met Leu Asn 40 Asp Thr Leu His Asn Ile Arg Thr Thr Asn Gln Ala Leu Lys Lys Glu 55 60 Leu Ser Gln Lys Thr Leu Thr Lys Thr Ser Leu Glu Glu Ile Ala Leu His Ser Ser Gln Ile Ser Met Asp Val Asn Lys Ser Ala Gln Leu Leu 90 85 Asp Ile Leu Ser Arg Asn Glu Tyr Pro Ile Asn Lys Asp Ala Arg Glu 105 110 100 Leu Leu His Ser Ala Pro Lys Glu Ala Glu Leu Asp Gly Asp Gln Met 120 125 Ile Ser His Arg Glu Leu Trp Ala Lys Ile Ala Asn Ser Ile Asn Asp 135 140 130 Ile Asn Glu Gln Tyr Leu Lys Val Tyr Glu His Ala Val Ser Ser Tyr 150 155 Thr Gln Met Tyr Gln Asp Phe Ser Ala Val Leu Ser Ser Leu Ala Gly 165 170 Trp Ile Ser Pro Gly Gly Asn Asp Gly Asn Ser Val Lys Leu Gln Val 185 180 Asn Ser Leu Lys Lys Ala Leu Glu Glu Leu Lys Glu Lys Tyr Lys Asp 200 205 Lys Pro Leu Tyr Pro Ala Asn Asn Thr Val Ser Gln Glu Gln Ala Asn 220 215 210 Lys Trp Leu Thr Glu Leu Gly Gly Thr Ile Gly Lys Val Ser Gln Lys 235 230 Asn Gly Gly Tyr Val Val Ser Ile Asn Met Thr Pro Ile Asp Asn Met 245 250 Leu Lys Ser Leu Asp Asn Leu Gly Gly Asn Gly Glu Val Val Leu Asp 265 260 Asn Ala Lys Tyr Gln Ala Trp Asn Ala Gly Phe Ser Ala Glu Asp Glu 280 285 Thr Met Lys Asn Asn Leu Gln Thr Leu Val Gln Lys Tyr Ser Asn Ala 300 295 Asn Ser Ile Phe Asp Asn Leu Val Lys Val Leu Ser Ser Thr Ile Ser 310 Ser Cys Thr Asp Thr Asp Lys Leu Phe Leu His Phe 325 330

# (2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 64 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

# (ii) MOLECULE TYPE: peptide

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Met Val Thr Ser Val Arg Thr Gln Pro Pro Val Ile Met Pro Gly Met

1 5 10 15

Gln Thr Glu Ile Lys Thr Gln Ala Thr Asn Leu Ala Ala Asn Leu Ser

20 25 30

Ala Val Arg Glu Ser Ala Thr Ala Thr Leu Ser Gly Glu Ile Lys Gly 40 Pro Gln Leu Glu Asp Phe Pro Ala Leu Ile Lys Gln Ala Ser Leu Asp 55 50

- (2) INFORMATION FOR SEQ ID NO:47:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs

  - (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TAGACGACTA TAGCTCTTGC T

21